# Journal of Applied Psychology

Edited by Donald G. Paterson, University of Minnesota

Consulting Editors

George K. Bennett, *Psychological Corporation*
Harold E. Burtt, *Ohio State University*
Allen L. Edwards, *University of Washington*
Clifford E. Jurgensen, *Minneapolis Gas Co.*
Irving Lorge, *T. C. Columbia University*
Quinn McNemar, *Stanford University*
Alexander Mintz, *City College of New York*
James P. Porter, *Danville, Illinois*
Harold F. Rothe, *Amer. Hosp. Supply Corp., Evanston, Ill.*
Julian B. Rotter, *Ohio State University*
Edward K. Strong, Jr., *Stanford University*
Donald E. Super, *T. C. Columbia University*
Morris S. Viteles, *University of Pennsylvania*
Alfred C. Welch, *Knox-Reeves, Minneapolis*

## Table of Contents

*Unfair Employment Practices as Viewed by Private Employment Counselors:* V. Keenan and W. A. Kerr ..... 361
*The Accuracy of Application Blank Work Histories:* J. N. Mosel and L. W. Cozan ..... 365
*The Relationship between the Predictive Power of Aptitude Tests for Trainability and for Job Proficiency:* C. W. Brown and E. E. Ghiselli ..... 370
*Validation of a Clinical Approach to the Placement of Engineers:* R. Harrison and T. A. Jackson ..... 373
*Relationship of Masculinity-Feminity to Tests of Mechanical and Clerical Abilities:* M. L. Lee ..... 377
*Ability Patterns in Technical Training Criteria:* B. Fruchter ..... 381
*A Psychological Study of Occupational Adjustment:* A. Heron ..... 385
*How Supervise? Scores Before and After Courses in Psychology:* F. R. Wickert ..... 388
*Measurement of Supervisory Ability:* G. C. Carter ..... 393
*Studies of Group Behavior: Factors Associated with the Productivity of Groups:* J. G. Darley, N. Gross, and W. C. Martin ..... 396
*MMPI Personality Patterns for Various College Major Groups:* R. D. Norman and M. Redlo ..... 404
*The Effects of Tachistoscopic Training in an Adult Reading Program:* G. Manolakes ..... 410
*The Effect of Fluorescent Flicker on Visual Efficiency:* A. Zaccaria, Jr., and M. E. Bitterman ..... 413
*Visual Tracking: II. Effects of Brightness and Width of Target:* R. S. Lincoln and K. U. Smith ..... 417
*Comment on Darley's "Special Review":* K. Eells ..... 422
*Reply to Eells' Comment on Darley's "Special Review":* J. G. Darley ..... 424
*Book Reviews* ..... 425
*New Books, Monographs, and Pamphlets* ..... 433

American Psychological Association

**Vol. 36, No. 6** **December, 1952**

# Journal of Applied Psychology

Published Bi-monthly by the American Psychological Association, Inc.
Prince and Lemon Sts., Lancaster, Pa.

Annual subscription, $6.00; single copies, $1.25

*Subscriptions and business communications should be sent to*

American Psychological Association
1333 Sixteenth Street N.W.
Washington 6, D. C.

*Articles for publication and books for review should be sent to the Editor*

Professor Donald G. Paterson, Department of Psychology
University of Minnesota, Minneapolis 14, Minnesota

---

This Journal gives prompt consideration to manuscripts reporting *original investigations* in any field of applied psychology except clinical and consulting psychology. A descriptive or theoretical article is occasionally accepted if it deals in a distinctive manner with a problem of applied psychology. The policy is, however, to favor papers dealing with quantitative investigations of direct value to psychologists working in the following fields: Vocational diagnosis and occupational guidance; educational diagnosis, prediction and guidance at the secondary school level and higher; personnel selection, training, placement, transfer and promotion in business, industry and government service including the armed forces; supervisory training in business, industry and government; bio-mechanics or design of machines to fit the human operator; illumination, ventilation and fatigue in industry; job analysis, description, classification and evaluation; measurement of morale of executives, supervisors, or employees; surveys of opinion on social or political issues, such as those conducted by The Psychological Corporation; psychological problems in market research and in advertising.

Articles may be under 500 words. The maximum is 12,000 words, the average in the neighborhood of 4,000 words. To reduce lag of publication, adherence to the rule of "brevity consistent with clarity" is encouraged.

A lapse of six to twelve months occurs between acceptance of an article and its publication, the lag varying with the rate at which manuscripts are submitted. If, however, an author is prepared to defray the costs of printing the necessary extra pages, he may arrange for earlier publication without thereby postponing the appearance of manuscripts by other contributors. This enables the management to provide space in addition to the scheduled 64 pages per issue. "Early publication" is thus a direct contribution to the subscribers. By cutting down lag in publication, it also benefits those authors whose articles are published in regular turn.

Tables, footnotes and references as well as text of manuscripts should be typed double-spaced throughout. Authors should adhere to the conventions described in the "Publication Manual of the American Psychological Association," *Psychol. Bull.*, 1952, **49**, No. 4, Part 2. A copy of the Manual will be loaned to any prospective contributor who does not find it in his library.

Entered as second-class matter, August 19, 1943, at the post office at Lancaster, Pa., under the act of March 3, 1879

Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized October 10, 1947

Copyright, 1952, by The American Psychological Association, Inc.

Journal of Applied Psychology

VOL. 36, NO. 6 DECEMBER, 1952

# Unfair Employment Practices As Viewed by Private Employment Counselors *

Vernon Keenan and Willard A. Kerr

*Illinois Institute of Technology*

American business and industrial leaders make rather free use of the terms "freedom of opportunity" and "free enterprise." These terms have psychological meaning and significance; apparently they are generally accepted as integral parts of the American way of life. Yet, as in all cultures, there are dissonant elements and contradictions. Sometimes the contradictions seriously weaken the working strength of a culture. It is possible that unfair employment practices constitute a major inconsistency in the current American pattern. "Freedom of opportunity" has many implications, and one of the most important is equal opportunity to compete for jobs.

Psychologists should be interested in this problem. They alone among the scientists specialize in construction and use of the objective instruments which are the obvious alternatives to consideration of race, creed, and ideology in the employment process. Their own job opportunities and opportunities for service to society are, in fact, restricted according to the extent to which selection and placement testing is *not* institutionalized as a part of the industrial culture. This would seem to place the *real* interests of psychologists in fair employment practices at a level of professional self-interest and, therefore, not merely academic or political.

Examination of the psychological literature on the problem shows it to be either extremely sparse or only distantly relevant. Aside from the literature on attitudes and prejudice in social psychology (**1, 2**), the authors found few references specifically on the problem of adult employment discrimination. One closely related research by Saenger and Gordon (**4**) on 504 representative residents of New York City indicated lower job satisfaction on the part of minority group members and still lower job satisfaction on the part of individuals who had personally experienced discrimination.

## Employment Counselors

In speculating as to which occupational group has most insight into discrimination in employment, it was finally decided that counselors in private employment agencies were in the most strategic positions to obtain and report such confidential information. This seemed true because they handle primarily "the white collar" jobs, receive rather specific job orders from employers, and typically have a wide acquaintance with company employment officers.

*The Sample.* Accordingly, a random sample of private employment counselors was drawn. The State of Illinois requires that each such counselor register annually with its Department of Labor. The current list for the Chicago area contained 738 counselors. Every third name was drawn for an invitational sample, making a total of 246. Questionnaires were mailed to these 246 counselors, but 45 were returned unclaimed by the Postal Department. Of the remaining 201, 45 completed questionnaires were returned (22 per cent). Although anonymous, information data show the obtained sample to be representative of the total list as to sex distribution (51 per cent male) and age (means: male 33,

* This report was prepared from data gathered by means of a questionnaire constructed by Vernon Keenan while a student at the Illinois Institute of Technology.

female 26). Average experience was approximately five years. Aside from these facts, it was not possible to make a more refined check on the representativeness of those responding.

*The Questionnaire.*[1] The two-page questionnaire was preceded by a cover letter which stated that "As an employment counselor you have had an excellent opportunity to observe discrimination in hiring practices, and because of this we are asking you to assist us in determining the *extent* of such discrimination. We are not interested in whether such practices are proper or improper. As much as you can, then, try to answer the questions as objectively as possible. Your frankness and cooperation in this survey will be deeply appreciated. After you have completed the questionnaire, please return it in the stamped envelope which is enclosed. DO NOT SIGN THE QUESTIONNAIRE. If you desire a report on the results of this survey, print your name and address on the enclosed postal card and mail it *separately*."

The questionnaire itself obtained information on age, sex, and counseling experience. It also inquired into the predominant occupational level of job orders received, typical education of clients, and various aspects of discrimination.

## Results

*Ethnic Discrimination.* As shown in Table 1 wide differences are reported to exist in the difficulty of placing typical individuals of various ethnic groups even when all are equally qualified. Most difficult to place are Negroes. Next come Mexicans and Orientals. Of intermediate difficulty in placement are persons of Italian, Polish, and Slavic descent. Easiest to place are persons of Irish, English, and German descent. The range of difficulty in placement is suggested by the comparative facts that four-fifths of the counselors report Irish, English, and German applicants easy to place, whereas only two per cent report Negroes, Mexicans, and Orientals as easy to place. Eighty-four per cent of the counselors reported Negroes as "hard to place" or "never place" even if qualified.

Table 1

"How difficult is it to place persons of _____ descent, assuming that they are qualified?" *

| Ethnic Group | Always Place | Easily Place | Average Difficulty | Hard to Place | Never Place |
|---|---|---|---|---|---|
| 1. Irish | 31 | 51 | 13 | — | — |
| 2. English | 38 | 47 | 9 | 4 | — |
| 3. German | 33 | 47 | 16 | 2 | — |
| 4. Slavic | 4 | 33 | 51 | 7 | — |
| 5. Polish | 4 | 33 | 49 | 7 | — |
| 6. Italian | 2 | 38 | 38 | 20 | — |
| 7. Oriental | — | 2 | 22 | 62 | 4 |
| 8. Mexican | — | 2 | 22 | 56 | 11 |
| 9. Negro | — | 2 | 2 | 44 | 40 |

* Frequencies of responses are reported as percentages. Their sums horizontally are not equal to 100 because of the instances of "no reply."

The date of Table 1 resemble somewhat the Bogardus experience with "social distance" data; however, a slightly simpler explanation becomes apparent upon examination of the critical ratios of successive ethnic group ranks. The three easily placed groups (Irish, English, and German) are not significantly different among themselves in ease of placement according to these data. A second-level cluster (Slavic, Polish, Italian) of ethnic groups likewise do not differ significantly from each other in ease of placement. This is also true of the third level (Oriental, Mexican). However, each group in any level does differ significantly in ease of placement from any group in any other level including the fourth level (Negro). The homogeneity of variance and over-all variance ratio were established as significant before critical ratios were determined.

*Religious Discrimination.* The per cent of counselors reporting major religious groups as "always place" or "easily place" if qualified were as follows: Protestant, 84; Catholic, 75; Jewish, 9. No counselor reported either Protestants or Catholics as "hard to place" or "never place" even if qualified, but 46 per cent stated this to be true of Jews. It seems from

[1] The questionnaire has been filed with the American Documentation Institute. Order Document 3709 from American Documentation Institute, 1719 N Street, N.W., Washington 6, D. C., remitting $1.00 for microfilm (images 1 inch high on standard 35 mm. motion picture film) or $1.00 for photocopies (6×8 inches) readable without optical aid.

these data, at least in the Chicago area, that Jews are as difficult to place as are Orientals.

*Job Requirements.* The counselors were asked "When arranging an interview for a client, how frequently are you asked to supply the following information?" The items of information follow with the per cent of counselors reporting "always" or "most of the time": experience, 98; full name, 88; age, 84; personal appearance, 60; racial descent, 40; nationality, 36; address, 26; religion, 25.

They also were asked "When discussing the rejection of an applicant with an employer, how frequently are the following mentioned as reasons for the rejection?" The reasons follow with the per cent of counselors stating either "always" or "most of the time": experience, 91; age, 22; appearance, 13; race, 8; religion, 4; nationality, 4; address, 2; name, 0.

It appears from the first answers above that discrimination is specifically present in at least one-third of the job orders received by the counselors. However, this may underestimate the problem. One counselor appended the following typical note to her questionnaire. "The job order contains two blanks—Non-Nordic (Jew-Neg-Orien) & Nordic (Christians) thus enabling our counselors to prevent the embarrassment of sending a Jew, etc. to a Nordic company. So the nationality-religion problem never enters as a reason for rejection."

The finesse with which discrimination is handled is illustrated by the ambiguous use of the term "personality" as shown by the statement of another counselor. "It is taken for granted that in screening applicants for employment we will not recommend one whose personality or appearance is not acceptable. Experience gives counselors knowledge of types of applicants acceptable to company's representative, and company's representative a confidence in counselor's recommendations."

*Size and Extensity of Firm.* When asked "In your opinion, which of the following is most particular about the race or religion of job applicants?" the counselor respondents replied as follows: small firms, 36; large firms, 33; medium-sized firms, 18; no reply, 13. The same type of judgment was requested for local and national organizations. "In your opinion, which of the following is more particular about the race or religion of job applicants?" The per cents of respondents replying were as follows: local, 47; nation-wide, 38; no reply, 16. These results are difficult to interpret, but are believed not to indicate any extreme uniqueness of discrimination as related to size and extensity of firm.

*Policy Origin.* Who sets discriminatory policies? Although this study could not obtain detailed information on this question, the following aspects were investigated. "In your opinion, is selection by race and religion company policy, the personal policy of the personnel director, or the expressed wish of the employees?" The per cents of counselors giving each cause are as follows: company policy, 53; personnel director, 22; employees, 18; no reply, 7.

*Effect on Counselor.* "Do you feel that any discrimination that does exist makes your job less difficult, more difficult, or about the same?" The distribution of replies was as follows: more difficult, 67; about the same, 31; less difficult, 2. The older and more experienced counselors were more prone to reply "about the same" to this question than were the younger counselors who more consistently replied "more difficult." This is interesting, because replies to no other question were significantly related to age.

## Summary and Conclusions

A sample of private employment agency counselors in the Chicago area was requested by mail to cooperate in a study of unfair employment practices (22 per cent cooperated). The replies of the 45 respondents suggest the following tentative conclusions.

1. According to the reports of these employment counselors, the ethnic group which bears the severest brunt of job discrimination is the Negro; 84 per cent of the counselors report extreme difficulty in their placement "even if qualified."

2. The second cluster of ethnic groups in order of job discrimination experience includes the Mexicans and Orientals.

3. The third cluster (less intense discrimination) includes Slavs, Poles, and Italians.

4. The fourth cluster includes Irish, English, and Germans.

5. Religious discrimination is directed largely against the Jews.

6. Discrimination is reported as specifically apparent in at least one-third of the job orders received by the private agency counselors.

7. A majority of counselors responding believe unfair employment practices to be deliberate company policy.

8. Two-thirds of the respondents report that the "discrimination that does exist" makes their jobs more difficult.

9. The hypothesis that private employment agency counselors can supply critical information about the operation of discrimination in employment appears, from these results, to be verified.

*Received January 16, 1952.*

## References

1. Krech, D., and Crutchfield, R. S. *Theory and problems of social psychology.* New York: McGraw-Hill Book Co., 1948.
2. Murphy, G., Murphy, L. B., and Newcomb, T. M. *Experimental social psychology.* New York: Harper & Bros., 1937.
3. Nudell, I. G., and Paterson, D. G. Attitudes of clerical workers toward three types of employment agencies. *Personnel*, 1950, **26**, 330–334.
4. Saenger, G., and Gordon, N. S. The influence of discrimination on minority group members in its relation to attempts to combat discrimination. *J. soc. Psychol.*, 1950, **31**, 95–120.

# The Accuracy of Application Blank Work Histories

**James N. Mosel**

*The George Washington University*

**and**

**Lee W. Cozan**

*The Hechinger Company, Washington, D. C.*

Although work histories have long been a staple item in employment procedures, the accuracy with which such data are reported by the applicant has received remarkably little systematic quantitative investigation. The only adequate study is the recent work of Keating, Paterson, and Stone (**4**). These investigators studied the accuracy of three work history items (wages, length of previous employment, and job duties) obtained by counseling interviews. By correlating worker claims with records of previous employers, it was found that applicant reports were surprisingly accurate for both recent and remote employment, the correlations ranging from .90 to .98.

The most usual method, however, for obtaining data on previous employment is through a written questionnaire—the application blank in industry and Form 57 or "unassembled examination" in government. While the literature of personnel psychology contains numerous studies of the utility of such data to predict job success, the accuracy of the applicant's responses has apparently received little attention. It would appear hazardous to project the findings of Keating *et al.* to the application blank inasmuch as these results were obtained by interview and in a counseling situation where there was presumably little incentive to distort.

A survey of the literature reveals for the most part only "incidents" rather than systematic quantitative investigation of this question. The cases mentioned by Moore (**6**) are conventionally cited in this connection. In the army during World War I, it was found that only six per cent of those claiming to be skilled in a particular trade actually had adequate skill, while over 30 per cent proved totally inexperienced ("trade bluffers"). The U. S. Civil Service Commission reported that over 0.3 per cent of the applicants in one year were barred for attempting to deceive the Commission. Among applicants for a position in a city department in New York, two per cent were found to have criminal records although they had sworn in their applications that they had none.

More recently, Lipsett (**5**) reports that in the Second Civil Service Region, 46 per cent of the cases subjected to a field investigation revealed falsification in the application for employment. In an unpublished study, Goheen and Mosel (**2**) found that of 69 civil service applicants for the job of traffic and transportation clerk, about ten per cent submitted inaccurate statements concerning previous job duties. Seven per cent of these were favorable to the applicant, while only three per cent involved misstatements of major importance.

With military personnel, however, Harris (**3**) found that only rarely do individuals falsify on certain items pertaining to previous psychiatric experiences. In fact, he found a tendency to over-report such occurrences. As a consequence, Conrad and Ellis (**1**) suggest that military personnel may provide personal history information with greater accuracy than civilians ordinarily do.

Parten (**7**) has reviewed the scant literature on the test-retest reliability and the accuracy (as judged by objective records) of such personal data as age, car and telephone ownership, etc., as obtained in public opinion and market research. Correlations were in the vicinity of .90 and percentages of agreement were generally in the 80's.

The significance of the accuracy of application blank work histories is twofold. In the

first place, accuracy of response imposes a limit upon the potential utility of such data in predicting job success (assuming that there is no correlation between job success and the direction of distortion). Secondly, the amount of distortion would be an important determinant of the necessity of a second selection aid—the recommendation questionnaire or reference check. The primary (and perhaps only) value of this device is that it provides verification of the assertions in the application blank. Verification may contribute to the predictive validity of the application blank by providing a more adequate basis from which predictions could be made. Furthermore, the amount of distortion uncovered may in itself serve as a revealing consideration in applicant evaluation.

The purpose of the present investigation is to present some much needed evidence on the accuracy of work history data obtained by the application blank in a setting where distortion might be expected to occur, i.e., the usual industrial employment situation. In conducting such a study, it is essential to determine not only the *frequency* of errors but also the *magnitude*. To this end, the technique of correlational analysis employed by Keating *et al.* was used.

## Procedure

The application blanks studied were those submitted over a one-year period to the personnel department of a building supply and engineering company. These applications were for sales and office positions and included the usual information on nature of previous experience, dates of employment, weekly salary, and the names and addresses of previous employers for each job held. The applicants were not informed that their statements would be subjected to verification, although it must be assumed that at least some expected a reference check to be made.

Three items of work history were submitted to verification: weekly salary, duration of employment, and job duties. Verification was achieved in a few cases of a local nature by telephone check, but the majority required a mailed questionnaire or recommendation form. The return rate on this questionnaire was approximately 95 per cent. Further checks were made in many cases through a credit bureau. In making these inquiries, the former employer was not informed of the applicant's claims; rather he was requested to submit the desired information without such knowledge. This was done to prevent the former employer from "verifying" applicant claims without actually

Table 1

Relationship between Claimed and Verified Weekly Wage for 61 Male Applicants on Jobs Held 0–12 Months Prior to Application for Employment

$r = .93$

| Claimed Wage | Verified Weekly Wage | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $10–19 | 20–29 | 30–39 | 40–49 | 50–59 | 60–69 | 70–79 | 80–89 | 90–99 | 100 and Over | Total |
| 100+ | | | | | | | | | 1 | 1 | 2 |
| 90–99 | | | | | | | 1 | | 2 | | 3 |
| 80–89 | | | | | | | | 5 | | | 5 |
| 70–79 | | | | | | 1 | 10 | 1 | | | 12 |
| 60–69 | | | | | 3 | 10 | | | | | 13 |
| 50–59 | | | 1 | 1 | 12 | | | | | | 14 |
| 40–49 | | | | 6 | | | | | | | 6 |
| 30–39 | | | 4 | | | | | | | | 4 |
| 20–29 | | | | | | | | | | | 0 |
| $10–19 | 2 | | | | | | | | | | 2 |
| Total | 2 | 0 | 5 | 7 | 15 | 11 | 11 | 6 | 3 | 1 | 61 |

Table 2

Relationship between Claimed and Verified Weekly Wage for 65 Female Applicants on Jobs Held 0–12 Months Prior to Application for Employment

$r = .94$

| Claimed Wage | Verified Weekly Wage | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | $20–29 | 30–39 | 40–49 | 50–59 | 60–69 | 70–79 | 80–89 | Total |
| 80–89 | | | | | | | 1 | 1 |
| 70–79 | | | | | | 2 | | 2 |
| 60–69 | | | | 3 | 5 | | | 8 |
| 50–59 | | 1 | | 8 | | | | 9 |
| 40–49 | | 4 | 16 | | | | | 20 |
| 30–39 | 2 | 19 | | | | | | 21 |
| $20–29 | 3 | 1 | | | | | | 4 |
| Total | 5 | 25 | 16 | 11 | 5 | 2 | 1 | 65 |

checking the records and from "giving the applicant a break" by refusing to disclose falsifications.

Data on 61 men and 65 women were sufficiently complete to permit analysis. In order to investigate the possibility that memory distortion might operate differentially in terms of the recency with which the jobs were held, accuracy was studied separately according to remoteness in time. The time interval categories were identical with those adopted by Keating *et al.*, i.e., a 0–12 month category which included all jobs terminated one year prior to applying for the present employment; a 13–24 months category which included all jobs terminated from 13 months to two years before applications; and so on. Work histories of male and female applicants were analyzed separately to uncover any sex differences.

## Results

Data were analyzed and displayed in the same manner as used by Keating *et al.* in order to afford a comparison of results.

*One-Year Time Interval.* A scattergram of the relationship between claimed and verified weekly wage on jobs held 0–12 months prior to application for employment is shown in Table 1 for males and in Table 2 for females. The correlation is .93 for men and .94 for women.

The scattergrams for duration of employment are shown in Tables 3 and 4. The cor-

Table 3

Relationship between Claimed and Verified Duration of Employment for 61 Male Applicants on Jobs Held 0–12 Months Prior to Application for Employment

$r = .98$

| Claimed Duration (in months) | Verified Duration (in months) | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0–9 | 10–19 | 20–29 | 30–39 | 40–49 | 50 and Over | Total |
| 50 and over | | | 1 | | | 7 | 8 |
| 40–49 | | | | | 3 | | 3 |
| 30–39 | | | | 5 | | | 5 |
| 20–29 | | | 10 | | | | 10 |
| 10–19 | 1 | 10 | | | | | 11 |
| 0–9 | 24 | | | | | | 24 |
| Total | 25 | 10 | 11 | 5 | 3 | 7 | 61 |

Table 4

Relationship between Claimed and Verified Duration of Employment for 65 Female Applicants on Jobs Held 0–12 Months Prior to Application for Employment

$r = .98$

| Claimed Duration (in months) | Verified Duration (in months) | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0–9 | 10–19 | 20–29 | 30–39 | 40–49 | 50 and Over | Total |
| 50 and over | | | | | | 13 | 13 |
| 40–49 | | | | 2 | 2 | | 4 |
| 30–39 | | | | 2 | | | 2 |
| 20–29 | | | 2 | | | | 2 |
| 10–19 | | 12 | | | | | 12 |
| 0–9 | 32 | | | | | | 32 |
| Total | 32 | 12 | 2 | 4 | 2 | 13 | 65 |

relation between claimed and verified duration is .98 for both male and female applicants.

Not shown graphically is the relationship between claimed and verified job duties. Agreement was found in 95 per cent of the male cases and in 94 per cent of the female cases.

*Longer Time Intervals.* The correlation between claimed and verified wages for all jobs held over 12 months prior to application for employment was .91 for the male cases and .90 for the female cases. Table 5 shows the correlations for each specific time interval category within this longer period. Because the number of cases in each interval is so small, considerable reservation must be held in using these coefficients to estimate the correlations in the population. It is clear, however, that in the present sample, there is no indication of a relationship between accuracy of response and the recency of the employment reported, nor of any sex difference.

The correlation between claimed and verified duration of employment for the longer time interval was .97 for male applicants and .87 for the female applicants. Correlation coefficients for each specific time interval are shown in Table 6. Here again there appears no evidence of decreased accuracy for the less recent jobs, nor is there any indication of a sex difference.

As for job duties, agreement between applicant claim and employer report was found in 84 per cent of the male applicants and in 87 per cent of the female applicants.

Table 5

Correlations between Claimed and Verified Weekly Wage for Male and Female Applicants on Jobs Held at Time Intervals Greater than One Year Prior to Application for Employment

| Time Interval (Job termination date prior to application for employment) | Men | | Women | |
|---|---|---|---|---|
| | N | $r$ | N | $r$ |
| 13–24 | 21 | .98 | 18 | .94 |
| 25–36 | 8 | .95 | 10 | .98 |
| 37–48 | 7 | .96 | 10 | .89 |
| 49–60 | 4 | .97 | 7 | .95 |
| 61 and over | 8 | .90 | 7 | .90 |

Table 6

Correlations between Claimed and Verified Duration of Employment for Male and Female Applicants on Jobs Held at Time Intervals Greater than One Year Prior to Application for Employment

| Time Interval (Job termination date prior to application for employment) | Men | | Women | |
|---|---|---|---|---|
| | N | $r$ | N | $r$ |
| 13–24 | 21 | .99 | 18 | .91 |
| 25–36 | 8 | .99 | 10 | .99 |
| 37–48 | 7 | .97 | 10 | .99 |
| 49–60 | 4 | .99 | 7 | .95 |
| 61 and over | 8 | .98 | 7 | .98 |

Inspection of the above tables brings out two additional observations: 1. Of the relatively few distortions that do occur, the great majority are in the direction of overestimation and hence in the applicant's favor. This was not apparent in the data of Keating *et al.* 2. For the more recent employment, wage claims are somewhat more susceptible to distortion than are claims concerning duration of employment. This difference was not apparent for the more remote jobs. These same relationships also hold for the data of Keating *et al.* The fact that items have a differential susceptibility to distortion is further supported by the study by Goheen and Mosel (**2**) on traffic and transportation clerks. Ten per cent of these applicants gave advantageously inaccurate replies to the item "reasons for separation"; seven per cent on job duties; and one per cent on wages.

## Summary

The present study was undertaken to determine the accuracy of work history data as contained in the applications for employment submitted to a building supply and engineering company. High agreement was found between applicant claims and verifications by past employers with respect to weekly wages, duration of employment, and job duties. All correlations except one were .90 or greater. Furthermore, there was no evidence of a relationship between validity of claim and recency of job; nor was there any apparent sex difference.

There was, however, a difference in the susceptibility to distortion of items for the more recent jobs. Wages were more subject to misstatement than was duration of employment.

These results are in very close agreement with the findings of another study in which the same items were obtained by counseling interview. It might have been expected *a priori* that the employment situation of the present study would have encouraged greater distortion than the counseling interview. This, however, does not seem to be the case. Apparently, the implied verification of application data in the employment situation is a sufficient deterrent to distortion to provide an accuracy comparable to that which the counseling situation achieves through lack of incentive to distort. One point of difference, however, was noted. In the present study distortions tended markedly in the direction of overestimation, a trend not found in the study of the counseling interview. It may well be that the two situations differ not so much in *amount* of work history distortion as in the *direction* of distortion.

*Received November 27, 1951.*

## References

1. Conrad, H. S., and Ellis, A. The validity of personality inventories in military practice. *Psychol. Bull.*, 1948, **45**, 385–426.
2. Goheen, H. W., and Mosel, J. N. The accuracy of applications for Civil Service employment. Unpublished study, 1950.
3. Harris, H. J. The Cornell Selectee Index—an aid in psychiatric diagnosis. *Annals N. Y. Acad. Sci.*, 1946, **46**, 594–603.
4. Keating, Elizabeth, Paterson, D. G., and Stone, C. H. Validity of work histories obtained by interview. *J. appl. Psychol.*, 1950, **34**, 1–5.
5. Lipsett, L. The personal investigation in the selection of employees. *Pers. Admin.*, 1946, **9**, 23–29.
6. Moore, H. *Psychology for business and industry*. New York: McGraw-Hill, 1942, p. 71.
7. Parten, Mildred. *Surveys, polls and samples*. New York: Harper, 1950, Chap. 16.

# The Relationship between the Predictive Power of Aptitude Tests for Trainability and for Job Proficiency

**C. W. Brown and E. E. Ghiselli**

*University of California, Berkeley*

The success of workers has been measured in a variety of ways, but it is possible to distinguish at least two broad categories of criteria, namely, success in training and degree of job proficiency achieved after training. In many instances it is difficult to clearly differentiate between these two types of criteria. For example, on some jobs the worker receives his instruction quite informally, and if his success is measured in terms of production any differentiation between the learning period and the post-learning period is purely arbitrary. Furthermore, for certain jobs longitudinal studies of worker proficiency indicate continued improvement over a period of years and, therefore, it is hard to say just where the learning period terminates.

Nevertheless, it is often feasible to make a distinction between trainability and level of job proficiency as measures of worker success, and ordinarily such a differentiation is made. Examples of training criteria are grades earned in formal training courses and instructors' or supervisors' ratings of worker progress during the learning period. Job proficiency has been measured in a variety of ways including speed and amount of production, achievement tests, supervisors' ratings, and the like.

Knowing that at least a rough distinction can be made between these two classes of criteria, we can consider a further problem. Only infrequently is attention paid to whether the types of abilities important in learning the skills and knowledges required for a job are the same as those that are important for performance after these skills and abilities have been acquired. Undoubtedly, few people would hold that they are identical, although most would expect considerable communality. The question arises: Can measures in one of these areas (learning or job proficiency) be used to predict success in the other area?

Obviously, the nature of the index utilized to measure success in learning and the index used to measure job proficiency may differ considerably. Thus, in the training of welders their improvement might be measured in terms of performance on work sample tests, while the success on the job itself might be gauged in terms of supervisors' ratings. In this case, an aptitude test which forecasts the learning criterion to a particular degree might be expected to forecast the job proficiency criterion to a somewhat different degree. Nevertheless, it would be expected that a test which had no validity for the one criterion would at best have low validity for the other, and a second test that had high validity for the one would be at least moderately valid for the other.

In other words, one would expect to find a positive correlation of at least moderate magnitude to hold between the validity coefficients of various tests determined on the basis of their predictive power for training and their predictive power for proficiency. Indeed, knowing that scores on a test are related to trainability of workers for a particular job to a particular degree one is very likely to generalize to the conclusion that it has approximately the same validity for predicting proficiency, and vice versa. The purpose of the present paper is to investigate this hypothesis that there is a substantial correspondence between the degree to which a test is valid in the prediction of trainability and the degree to which it is valid in predicting job proficiency.

## Methods and Procedure

The present writers have developed a file of investigations conducted in the United States since 1919 relative to the validity of tests. In this file tests and occupations are classified by type. It is possible, then, to obtain a summary of the findings concerning the validity a particular type of test has for a particular occupation either with respect to training or with respect to job proficiency. In summariz-

ing such a validity we have computed the weighted mean validity coefficient through Fisher's $z'$ transformation. We have considered the shortcomings of this procedure elsewhere.[1]

Our approach, then, is to take as the index of validity the average of the validity coefficients found in previous studies. Thus, ideally, for each type of test and for each type of job we would have the average validity coefficient in the prediction of trainability and the average coefficient in the prediction of job success. Unfortunately, the data available from previous investigations are incomplete and, therefore, we do not have a pair of validity coefficients for each type of test for each type of occupational group. Hence, the comparisons and relationships given in this report are not as complete as would be desirable.

The classification of tests which we have employed is given below. Detailed descriptions of the various rubrics are to be found in the previously mentioned report.

*Intellective:* Intelligence, Immediate memory, Substitution, and Arithmetic; *Speed of Perception:* Number comparison, Name comparison, Cancellation, and Speed of perception; *Spatial:* Pursuit, Location, Tracing, and Spatial relations; *Motor:* Tapping, Dotting, Finger dexterity, Hand dexterity, and Arm dexterity; and *Mechanical Principles.*

The categories into which all jobs were classified are as follows: *Clerical Occupations:* General clerks, Recording clerks, and Computing clerks; *Protective Service Occupations; Personal Service Occupations; Vehicle Operators; Trades and Crafts:* Mechanical repairmen, Electrical workers, Structural workers, Processing workers, Complex machine operators, and Machining workers; and *Manipulative and Observational Occupations:* Machine tenders, Bench workers and assemblers, Inspectors, Packers and wrappers, and Gross manual workers.

## Results

All data pertinent to the present problem are given in Table 1. The first question concerns the degree of relationship obtaining between the validity coefficients for the two types of criteria. Taking all tests and jobs together, a total of 127 pairs of validity coefficients were available. For these cases the coefficient of correlation was found to be .17. The second column of the table presents coefficients by type of test and type of job. It will be noted that for these various breakdowns the relationship is of about the same order. It is apparent from these data that the relationship between the validity coefficients of tests for trainability and their validity for job proficiency is low. There is only a slight tendency for a test that proves useful in predicting trainability also to be useful in predicting job proficiency, and vice versa.

Table 1

Relationships between Validity Coefficients of Tests Used in the Prediction of Trainability and of Job Proficiency, together with Median Differences between Coefficients

| | N | $r$ between Validity Coefficients | Median Difference between Coefficients |
|---|---|---|---|
| Type of Test | | | |
| Intellective | 35 | .21 | .11 |
| Speed of Perception | 19 | .14 | .13 |
| Spatial | 38 | .12 | .07 |
| Motor | 22 | .09 | .07 |
| Type of Job | | | |
| Clerical | 31 | .20 | .12 |
| Trades and Crafts | 68 | .15 | .11 |
| Manipulative and Observational | 21 | .22 | .10 |
| All Tests and Jobs | 127 | .17 | .10 |

In the final column of the table are given the medians of the differences between the validity coefficients for the two types of criteria. Taking all tests and jobs together the median difference between the coefficients is .10. While at first glance this difference may not be considered particularly large, it is to be remembered that the validity coefficients for separate tests are seldom found to be over .50. Therefore, in general, the median difference is about one-fifth of the effective range of validity coefficients. Reference to the median differences for breakdowns of the individual tests

[1] Ghiselli, E. E., and Brown, C. W. Validity of aptitude tests for predicting trainability of workers. *Personnel Psychol.*, 1951, 4, 243–260.

and jobs as given in Table 1 indicate that they do not vary significantly from the median value for the total set of data.

## Discussion

The findings herein presented indicate that only a low relationship exists between the validity of tests in the prediction of trainability and in the prediction of job proficiency. That is, a test which is found to have high validity in the prediction of workers' capacity to learn a job may have relatively little validity in predicting how well they will perform on that job after their training has been accomplished. These results suggest that the abilities important for learning a job may differ markedly from those important in the maintenance of proficiency on the job.

There are, of course, a number of laboratory investigations indicating that the abilities involved in the early stages of learning a task differ from those involved in later stages. It seems likely that a similar situation would exist with respect to an industrial task. However, it would be hazardous to draw this conclusion on the basis of the present analysis since the coefficients here utilized are subject to some distortion due to such factors as unreliability of criteria and restriction of range of talent. In any event it is clear that one cannot be at all secure in using a test that has been validated against training criteria as a basis for forecasting job proficiency, or in using a test validated against job proficiency criteria as a basis for forecasting success in training.

*Received December 6, 1951.*

# Validation of a Clinical Approach to the Placement of Engineers

**Ross Harrison*** **and Theodore A. Jackson**

*Stevenson, Jordon and Harrison, Inc., New York City*

While there have been innumerable validity studies which correlate batteries of psychological tests with job proficiency criteria and while there have also been many attempts to predict the achievement of students in engineering school from psychometric measures, the technical literature is singularly unrewarding on the subject of predicting job success of engineers in industry. The present note is a report on a personnel evaluation program with mechanical engineers in the Wright Aeronautical Division of the Curtiss-Wright Corporation.[1]

Usually articles on the validity of personnel selection procedures are primarily statistical studies which relate batteries of tests to measures or ratings of job adjustment. In recent years it is being realized in some quarters that a strictly psychometric or statistical approach to occupational prognosis is limited because it ignores psychological characteristics which, while not directly measurable, are nevertheless important and may be evaluated. The use of multiple correlational methods with professional, sales, and executive personnel is less rewarding because the tests that make up the batteries deal more with cognitive abilities and do not adequately represent personality dimensions. Furthermore, such an approach neglects not only intrapersonal dynamics but also the integration of segmental information about the individual which is necessary for understanding the probabilities of job success. The global or holistic methods of clinical psychology such as projective techniques and depth interviewing allow for greater synthesis of relevant data, whether qualitative or quantitative, and are coming into wider use in personnel selection by some psychologists working in the industrial area. The usual objection to the clinical approach is that it is not objective and cannot be properly validated and hence is not scientifically acceptable. The rejoinder is that, while it is true that clinically oriented psychologists seem less inclined to validate their procedures in a rigorous manner and while the clinical methods are admittedly subjective, the effect of applying these methods can be objectively and empirically validated. In a small way we hope, in this article, to contribute to the validation of clinical procedures in industrial placement. At the same time we realize that, because of practical considerations, our methods of validation are lacking in methodological refinement.

## Procedure

The population was made up mostly of recent graduate engineers who had not yet been assigned to stable positions within the Engineering Department. In the evaluation program each engineer was studied by means of a personal history form and a clinical interview as well as with a battery of ability, aptitude, interest, and personality tests. During the course of the interview, which was primarily diagnostic in nature, the test findings were communicated to the engineer for his information and guidance. Test evaluation was qualitative with every source of information considered in relation to all other relevant data. Emphasis was placed on patterned relationships and not on discrete scores or segmental behavior. Eventually a confidential written report, which included an analysis of abilities, interests, personal traits, placement recommendations, and developmental potentialities for each man, was submitted to the technical administrators who were responsible for assignment.

The evaluative techniques were as follows: Wonderlic Personnel; Shipley-Institute of Living Vocabulary and Abstract Reasoning; Otis Arithmetic Reasoning; Bennett Mechanical

* Hunter College.

[1] We are grateful to Messrs. G. R. Heil and P. R. Jackson of the Engineering Department for expediting the study. Mr. J. L. Altenhaus was responsible for most of the computations.

Comprehension (Form BB); DAT Space Relations; Strong Vocational Interest Blank; Thematic Apperception Test (8 picture abridgment); personal history form; personality questionnaire; and a projective verbal situations test. The last two tests and the personal history form were custom-made for our purposes and are not on the market. The Engineering and Physical Science Aptitude Test of Moore, Lapp, and Griffin was tried out on a small sample but had to be abandoned because it did not discriminate sufficiently for our highly selected population. Some of the other ability tests had ceilings which were short of ideal, but none was so low as to preclude fairly adequate discrimination for practical purposes.

Because many of the personality tests were not scoreable and because of the integrative nature of our procedures, orthodox correlational methods could not be employed. A further difficulty was that no comprehensive merit ratings over an extended period were available on the engineers. Hence, it was necessary to develop other, if less familiar, methods for testing out the validity of our approach.

*First method.* In the first phase of the validation the personnel evaluation reports of 113 mechanical engineers were compared with subsequent job performance. The immediate supervisors of the engineers, who with rare exceptions had not had occasion to consult the reports, were asked to compare the personal descriptions and predictions therein with actual job behavior as they knew it from daily contact. Criteria for selecting the reports to be validated were twofold: (1) the psychologist's recommendation must have been followed; and (2) the supervisor was required to have at least three months' close observation of the man being rated. A member of the Engineering Department decided whether the recommendations had been followed or not.

The supervisor, after reading the report in each case, was requested to check one answer for the following question:

"From your experience with this man, would you say that you agree with the descriptions and findings contained within the report? High degree of agreement ____; predominant agreement or mostly agree ____; slight agreement ____; no agreement ____."

*Second method.* A complementary approach was made to validation by seeking the opinions of higher echelon administrators who were responsible for placement. These were the directors and division managers who had had approximately a year and a half experience with the evaluation program. There were three main questions:

1. "In your experience with the personnel evaluation reports, to what extent have you found them generally helpful and accurate? Very helpful and accurate ____; fairly helpful and accurate ____; somewhat helpful and accurate ____; not at all helpful nor accurate ____."
2. "Compare the testing program with the placement procedures that existed before testing was initiated. Has the testing program added anything useful or improved in any way the placement of engineers? The testing program is: A marked improvement in procedure ____; a slight improvement ____; no improvement ____; would prefer to dispense with it and rely on trial and error alone ____."
3. "Did your attitude become more or less favorably disposed to the personnel evaluation reports as you became better acquainted with them and made use of them? More favorable ____; less favorable ____; unchanged ____."

In each instance comments were encouraged. In addition, there was a question about the approximate number of reports that each administrator had read.

## Results

*Recommendations.* Some interest revolved around the question of the degree to which directors and division managers followed placement recommendations contained in the evaluation reports. No useful purpose would be served if the psychologists were academically correct but were operating in a vacuum with their recommendations being largely ignored.

In 28 per cent of the cases the findings with regard to disposition were ambiguous. When the ambiguities were disregarded and consideration was given only to those cases where a fairly sharp affirmative or negative decision was possible, it was found that 78 per cent of the recommendations had been followed. Because of personnel requirements of different groups within the Engineering Department, it would not be reasonable to expect that recommendations could always be followed.

*First method.* Table 1 gives the results on the extent of agreement between the reports and the judgment of first-line supervisors based on their experience with the men who had been evaluated. Agreement was substantial and was greater than had been anticipated from a knowledge of possible attenuating factors. The criterion itself was subjective and hence subject to human error. Even when follow-up investigations were carried out in cases of evident contradiction, it was frequently hard to determine whether the error lay with the psychological evaluation or with the supervisor's judgment or was due to semantic problems of communication or to unusual environmental conditions which might obscure the working out of the personal qualities reported in the evaluation. There can be no doubt, however, that in some instances the evaluations themselves were in error.

*Second method.* A tabulation of the opinions of the directors and division managers is presented in Table 2. The results are favorable but less so than for the more direct method of validation.

When the administrators were asked to what degree they had found the reports helpful and accurate, about three-quarters answered in the first two (or favorable) categories. No one said that the reports were valueless.

On the second question, which compared the evaluation program with the placement procedures that existed before testing was initiated, practically all the respondents agreed that there was some improvement. They were evenly divided as to the amount of agreement—whether "marked" or "slight."

In response to the third question, which was concerned with attitudinal changes to the program, 64 per cent were more favorably impressed with added experience, 8 per cent became more negative, while 28 per cent reported no change. Of those who reported no change, many had favorable attitudes from the beginning.

Information was available on the number of reports that each administrator had read for assignment purposes. An attempt was made to find out if reaction to the program was related to familiarity with reports. Further statistical computations, which will not be reported here, were made to determine the relationship. The resulting figures showed that for the first two questions those administrators who were more familiar with evaluation reports were also more favorably disposed to psychological testing, but this same relationship did not hold for the third question.

The solicited comments of the directors and division managers were highly diversified and showed only one definite trend. The majority of the remarks were concerned with administrative aspects of the evaluation program rather than with the content of the reports themselves. Since these remarks were frequently critical, this fact served to attenuate the validity figures. Critical comments about the administrative use to which the reports were put somewhat confused the issue, since the basic problem was the validity of the

Table 1

Agreement between Evaluation Reports and Ratings by Supervisors of 113 Engineers

| Degree of Agreement | N | Per Cent |
|---|---|---|
| High degree of agreement | 45 | 39.8 |
| Predominant agreement or mostly agree | 58 | 51.3 |
| Slight agreement | 9 | 7.9 |
| No agreement | 1 | .9 |

Table 2

Percentage Distribution of Ratings by Directors and Division Managers on the Personnel Evaluation Program

| Question 1. (N = 24) | Per Cent |
|---|---|
| Very helpful and accurate | 16.5 |
| Fairly helpful and accurate | 58.5 |
| Somewhat helpful and accurate | 25 |
| Not at all helpful and accurate | 0 |
| **Question 2. (N = 21)** | |
| A marked improvement in procedure | 47.6 |
| A slight improvement in procedure | 47.6 |
| No improvement | 4.8 |
| **Question 3. (N = 25)** | |
| More favorable attitude | 64 |
| Less favorable | 8 |
| Unchanged | 28 |

evaluations, not the way in which they were used. Moreover, many of the administrative problems were a consequence of the shortage of engineers and therefore beyond control.

## Summary

A clinical approach to the evaluation and placement of engineers in the aeronautical industry is described. Validation was studied in two ways.

1. First-line supervisors rated the accuracy of psychological reports on the basis of their experience with the men who had been evaluated. Substantial agreement was obtained in over 90 per cent of the cases.

2. Higher echelon administrators, responsible for assignment, after a year and a half experience gave their opinions on the value of the program. The results were again favorable but were slightly less so because of administrative complications.

*Received August 4, 1952.*
*Early publication.*

# Relationship of Masculinity-Femininity to Tests of Mechanical and Clerical Abilities *

**Marilyn C. Lee**

*University of Illinois*

The primary purpose of this investigation was to study relationships between aptitude and personality variables with specific reference to mechanical ability, clerical ability, and masculinity-femininity. A secondary purpose was to explore the possibility of developing a disguised and relatively non-fakable measure of M–F from clerical and mechanical test materials.

The rationale for this study is based on consistent findings reported in the literature indicating that women tend to score higher than comparable groups of men on tests of clerical aptitude (**5**, **6**, **11**), while men perform better than women on most tests of mechanical ability (**3**, **10**, **13**). It should be emphasized that these ability differences are reported in reference to the biological dichotomy of males and females. Are such differences also reflected in the psychological continuum of masculinity-femininity as measured by an inventory of interests, attitudes, and modes of behavior?

## Tests and Subjects

The tests administered for the purposes of this study included the Bennett Test of Mechanical Comprehension, Form AA, the Minnesota Clerical Test, and the Terman-Miles Attitude-Interest Analysis Test, Form B. The Bennett and Minnesota tests were selected because they have repeatedly shown substantial sex differences, and are of an appropriate difficulty level for the sample utilized; the Terman-Miles inventory was used because it represents one of the most comprehensive and carefully constructed inventories yet developed for the measurement of masculinity-femininity. Since performance on these tests is related in varying degrees to performance on measures of general academic ability (**1**, **2**, **14**), available scores on the American Council on Education (ACE) Test for College Freshmen were also utilized in the analysis.

The experimental group comprised 102 male and 72 female students recruited from a sophomore level course in psychology at the University of Minnesota. Biographical information obtained for all subjects from questions prefacing the Terman-Miles test indicated that the sample was quite uniform with respect to regional background, the great majority being local residents. The mean ACE score of the group was at the 63rd percentile on national college freshman norms.

## Results

*Inter-Sex Differences.* Means and standard deviations of the four test variables are presented separately for the male, female, and combined groups in Table 1. These results corroborate previous research in showing that men as a group perform significantly better than women on the Bennett Test of Mechanical Comprehension (C.R. = 10.7), while women are significantly superior to men on the Minnesota Clerical Test (C.R. = 3.6). The sharp sex-differentiating power of the Bennett test is evidenced by the relatively small overlap (21.7%) of the male and female score distributions. Although the Minnesota Clerical Test also succeeds in dichotomizing the sexes, it is less efficient in this respect than the Bennett, as indicated by the larger score distribution overlap of 45.9%. The superior differentiating ability of the mechanical test may be at least partially attributable to its greater amenability to cultural influence.

By way of comparison, the Terman-Miles Attitude-Interest Analysis Test separated males and females with a percentage overlap of only 9.7, an amount comparable to the 8% average overlap obtained by the test authors for total scores on this test (**14**). Results for the ACE test bore out expectations based on previous research (**15**) that there would be no

* This article is based on the writer's M.A. thesis of the same title, completed in December, 1950 under the direction of Prof. Donald G. Paterson, and on file in the University of Minnesota Library.

Table 1

Means and Standard Deviations for the Four Test Variables

| Test | Males (N = 102) M | Males (N = 102) S.D. | Females (N = 72) M | Females (N = 72) S.D. | Total Group (N = 174) M | Total Group (N = 174) S.D. |
|---|---|---|---|---|---|---|
| Bennett Mech. Compr. Test, Form AA | 44.7 | 8.6 | 30.2 | 9.0 | 38.7 | 11.3 |
| Minn. Clerical Test | 244.2 | 46.9 | 270.5 | 46.8 | 255.1 | 48.6 |
| Terman-Miles Att.-Int. Anal. Test, Form B | +88.9 | 40.6 | −26.9 | 54.1 | +41.0 | 73.7 |
| ACE Psychol. Exam.* | 87.2 | 25.0 | 86.1 | 25.8 | 86.8 | 25.3 |

* Scores on the ACE test were available for approximately 80% of the experimental group. N's for this test are as follows: males, 84; females, 58; total group, 142.

statistically significant differences between the sex groups in performance.

Product moment correlations of the Terman-Miles test with each of the three other test variables are presented for the total group and for each sex separately in Table 2. In referring to this table it should be noted that plus scores on the Terman-Miles test indicate masculinity and minus scores indicate femininity. Thus, positive correlations between M-F and other test scores represent a positive relationship with masculinity and a negative relationship with femininity, the relationships being reversed for negative coefficients.

For the total sample, the correlation of .57 between the Bennett Mechanical and the Terman-Miles tests indicates that of the test variables employed the Bennett is the best single predictor of Terman-Miles M-F. Indeed, for this group the Bennett test predicts masculinity-femininity as well or better than it predicts various mechanical criteria (**4**, **8**, **9**, **12**), and correlates about as high with the Terman-Miles test as do many M-F tests with each other (**7**). Thus, if a mixed-sex group is used as the frame of reference, the Bennett may in a sense be referred to as a disguised measure of masculinity-femininity. Moreover, the Bennett predicts Terman-Miles M-F as well when used alone as when supplemented by the Minnesota Clerical Test. The obtained multiple R of .58 indicates that little predictive efficiency is gained by using both tests.

Table 2

Product Moment Correlations of the Terman-Miles Attitude-Interest Analysis Test with the Ability Test Variables

| Test Variables | Males | Females | Total Group |
|---|---|---|---|
| Terman-Miles M-F and Bennett Mech. Compr. | .11 | .24* | .57** |
| Terman-Miles M-F and Minn. Clerical | −.09 | .05 | −.22** |
| Terman-Miles M-F and ACE Psychol. Exam. | .08 | .26* | .12 |

* Significant at 5% level of confidence.
** Significant at 1% level of confidence.

The Terman-Miles M-F and Minnesota Clerical tests correlate −.22 with each other, indicating that for the sexes combined, measured femininity is positively related to clerical proficiency, but that the relationship is not nearly as high as that between masculinity and mechanical ability. Since the sex overlap of the Minnesota score distributions is more than twice that of the Bennett distributions, this result might well be expected. One of the reasons for the difference in correlation magnitude may be that the Bennett is more heavily loaded with experiential and informational factors than the Minnesota Clerical Test. Not only is the sex difference in means comparatively smaller for the Minnesota than for the Bennett, but the improvement that takes place with age in adolescence is also proportionately smaller for the Minnesota test: 9% as compared with a 23% improvement on the Bennett (**1**, **2**). Both of these findings suggest that experience affects performance to a greater degree on the mechanical than on the clerical test. If the posited information load-

ing does exist, then the Bennett test is probably in part a measure of interest as well as of aptitude. It should therefore be expected to correlate more highly with the M-F inventory, which also measures interest factors.

The obtained correlation of .12 between the ACE and the Terman-Miles M-F tests indicates that intelligence and masculinity-femininity are at best negligibly correlated in a mixed sample. This finding is consistent with the general trend of research which has shown little or no difference between comparable groups of males and females on standard tests of intelligence (**16**).

It should also be pointed out that although the mechanical and clerical tests both correlate positively with the ACE test, they correlate significantly negatively with each other (−.24). This negative correlation is probably due mainly to the opposite direction of sex differences, for if general ability is partialled out of both tests the negative correlation becomes, as anticipated, somewhat greater (−.35). However, the case of two tests which are related positively to general ability and negatively to each other is uncommon in the field of psychometrics, and suggests interesting applications of partial correlation and suppression techniques.

*Intra-Sex Differences.* The discussion thus far has been concerned with results obtained on a mixed sex sample. The inter-sex differences obtained for both the Bennett Mechanical and the Minnesota Clerical tests are substantial, and the Bennett test, in particular, is a significant predictor of Terman-Miles M-F scores. However, it was realized from the start that a more convincing demonstration of the predictive efficiency of ability test scores for the postulated trait of psychological masculinity-femininity could be obtained from correlations reported for each sex separately.

While the coefficients presented in Table 2 indicate that Terman-Miles M-F is not significantly associated with measured mechanical or clerical abilities within the male group, the obtained correlations are in the expected directions. For women, M-F is correlated to a low but statistically significant degree with both Bennett and ACE scores, but not with Minnesota Clerical scores. These results indicate that mechanical comprehension tends to be a masculine characteristic for women, and corroborate Terman and Miles' finding that brighter college women tend as a group to rank slightly more masculine than their less able colleagues (**14**).

The best prediction of Terman-Miles M-F for women is obtained by using both the Bennett and ACE tests. Although the obtained multiple R of .30 is not high enough for practical purposes, the data were derived from a sample which was very homogeneous with regard to age, education, and regional background.

In general, most of the within-sex differences are in the same direction as the obtained inter-sex differences but are much reduced in magnitude. These reductions in correlation size are much greater than would be expected from attenuation due to restriction of the total sample range. Such results suggest either: (1) that Bennett Mechanical and Minnesota Clerical test scores are not particularly good indicators of intra-sex M-F; or (2) that the validity of the Terman-Miles test as a measure of M-F within each sex should be more carefully investigated.

## Summary

The following major conclusions were derived from the data obtained in this investigation:

1. Men perform significantly better than women on the Bennett Test of Mechanical Comprehension, Form AA, and women demonstrate a somewhat less marked superiority on the Minnesota Clerical Test.

2. For a mixed-sex group, the Bennett Mechanical test predicts Terman-Miles M-F as well or better than it does most of the mechanical criteria against which it has been validated. Moreover, its correlation with the Terman-Miles test compares favorably with reported intercorrelations of various tests specifically designed to measure masculinity-femininity.

3. Although the Minnesota Clerical Test predicts Terman-Miles M-F for a mixed-sex group at a statistically significant level, it does not contribute substantially to the Bennett in multiple prediction of masculinity-femininity.

4. Most of the intra-sex correlations between ability test and M-F test scores are essentially in the same direction as those ob-

tained for the total group, but are much reduced in magnitude. These reductions in magnitude are in excess of what might be expected due to restriction of score range within single-sex groups.

5. The interpretation of the intra-sex correlations obtained in this study depends in part on the validity of the Terman-Miles test as a measure of psychological masculinity-femininity. It has been repeatedly demonstrated that the test is an extremely efficient discriminator of biological maleness-femaleness (**14**). In addition, Terman and Miles cite case histories of homosexuals who obtained scores characteristic of the opposite sex on the M-F scale (as well as high scores on the inversion scales). However, there is no conclusive evidence that the test is measuring intra-sex M-F with a high degree of validity within the normal range of the population. In order to develop such a test it would be necessary to employ the additional criterion of ratings of psychological M-F using a sufficient number of competent judges to insure adequate reliability. Terman and Miles used only a few judges and were discouraged by the low reliabilities which they obtained. However, by increasing the number of raters to 30 or 40 through the use of sorority or fraternity groups, it should not be too difficult to achieve satisfactory reliabilities, as can be readily estimated by application of the Spearman-Brown formula.

6. One of the original purposes of this project was to explore the feasibility of developing a disguised measure of masculinity-femininity from ability-type tests. As Terman and Miles (**14**) have pointed out, their M-F test is easily fakable if subjects are motivated to this end, even though the purpose of the test is otherwise difficult to discern. A non-fakable M-F test would be desirable, especially in selection situations, and if the test were of the ability type it would have the added virtue of providing both personality and ability scores for the same expenditure of time. Since it has a fairly high correlation with the Terman-Miles test within mixed-sex groups, and since it is so successful in differentiating the sexes, the Bennett Test of Mechanical Comprehension could conceivably serve as the basis for developing such a test. It could be constructed by analyzing a large pool of items similar to those found in the original Bennett and selecting those which most successfully differentiate men from women and at the same time serve as predictors of mechanical ability. As an additional criterion for item selection, trait ratings as proposed above should be used in order to insure high intra-sex differentiation.

*Received January 10, 1952.*

## References


1. Andrew, Dorothy M., and Paterson, D. G. *Manual for the Minnesota Clerical Test.* (Rev. Ed.) New York: Psy. Corp., 1946.
2. Bennett, G. K. *Manual for the Test of Mechanical Comprehension, Form AA.* New York: Psy. Corp., 1940.
3. Bennett, G. K., and Cruikshank, R. M. Sex differences in the understanding of mechanical problems. *J. appl. Psychol.*, 1942, **26**, 121–127.
4. Bennett, G. K., and Fear, R. A. Mechanical comprehension and dexterity. *Personnel J.*, 1943, **22**, 12–17.
5. Englehardt, Olga E. The Minnesota Clerical Test: sex differences and norms for college groups. *J. appl. Psychol.*, 1950, **34**, 412–414.
6. Goodenough, Florence L. The consistency of sex differences in mental traits of various ages. *Psychol. Rev.*, 1927, **34**, 440–462.
7. Heston, J. C. A comparison of four masculinity-femininity scales. *Educ. psychol. Measmt.*, 1948, **8**, 375–387.
8. McDaniel, J. W., and Reynolds, W. A. A study of the use of mechanical aptitude tests for selection of trainees for mechanical occupations. *Educ. psychol. Measmt.*, 1944, **4**, 191–197.
9. Moore, B. V. Analysis of tests administered to men in engineering defense training courses. *J. appl. Psychol.*, 1941, **25**, 619–635.
10. Paterson, D. G., Elliott, R. M., Anderson, L. D., Toops, H. A., and Heidbreder, E. *Minnesota mechanical ability tests.* Minneapolis: Univ. of Minn. Press, 1930.
11. Schneidler, Gertrude G., and Paterson, D. G. Sex differences in clerical aptitude. *J. educ. Psychol.*, 1942, **33**, 303–309.
12. Shuman, J. T. Value of aptitude tests for factory workers in the aircraft engine and propeller industries. *J. appl. Psychol.*, 1945, **29**, 156–160.
13. Stenquist, J. L. Measurements of mechanical ability. New York: *Teachers College Contributions to Education*, No. 130, 1923.
14. Terman, L. M., and Miles, Catherine C. *Sex and personality.* New York: McGraw-Hill, 1936.
15. Thurstone, L. L., Thurstone, Thelma G., and Adkins, Dorothy C. The 1938 Psychological Examination. *Educ. Rec.*, 1939, **20**, 263–300.
16. Tyler, Leona E. *The psychology of human differences.* New York: Appleton, 1947.

# Ability Patterns in Technical Training Criteria [1]

**Benjamin Fruchter**

*The University of Texas*

A number of different systems for classifying occupations into job families have been proposed (**3**, **10**, **12**, **13**). Several investigators (**1**, **8**) have applied factor analysis methods to the problem. The classifications have been based largely on the common elements such as skills, intelligence, and personality traits needed for successful performance in the occupations.

The present approach attempts to classify occupations on the basis of the profile of aptitudes required of the persons who are candidates for training. It should be particularly applicable to vocational and educational guidance and in the Armed Services, where a period of training usually precedes assignment to an occupation. Occupations which are quite different at the level of skilled performance may require similar profiles of aptitudes of applicants for training. Thus, it might be found that candidates for training as weather observers and clerk-typists require similar aptitudes although the jobs themselves are quite different above the training level.

Both Guilford (**5**) and Gulliksen (**7**) have stressed the importance of analyzing criteria from the factorial approach so that the relative contribution of various aptitudes and traits to successful performance might be evaluated. Scores on a battery of selection or classification tests can then be related to ultimate performance in one or more occupations, and appropriate assignment can be made.

## Population and Factor Analysis

The present exploratory study was done with data from the USAF Air Training Command. A battery of classification tests is routinely administered to airmen, and the results are used to assign them to technical schools for training in various occupational specialties. To evaluate the effectiveness of the tests in selecting successful candidates for training, the scores are validated against the final grades made in the technical schools.

The following two alternative approaches suggest themselves for determining the factor content of these final-grade criteria:

1. Consider each technical specialty a separate population and perform a factor analysis of the intercorrelations of its criterion and the test battery, or,
2. Consider the airmen a single population, perform a factor analysis of the battery on a representative sample. Estimate the factor content of each training criterion on the basis of the factor loadings obtained from the combined population and the validity of the tests.

The latter approach was adopted because the loadings from specialty to specialty would be more comparable and a prohibitive number of factor analyses would not be required.

A representative sample of 389 airmen who had been assigned to training in technical schools was drawn. Each school was represented in the sample in proportion to the total number of airmen who had been sent into that type of training, and the individuals chosen to represent each school were randomly selected.

The students were all male, mostly 18 and 19 years old, and had on the average a tenth-grade education. The intercorrelations of the 19 tests administered to them were obtained from the Quarterly Research Reports of the USAF Air Training Command, 3309th Research and Development Squadron (**11**, p. 5). Six factors were extracted by the centroid method and rotated to psychologically meaningful positions. Table 1 gives the rotated loadings for the 19 variables. The orthogonal rotated factors are identified as follows: I. Numerical facility (N); II. Verbal

[1] Based on a paper read at the American Psychological Association meetings in September, 1950. Although this study was done while the writer was a civilian employee of the Air Training Command, the views expressed in this article are those of the writer and do not necessarily represent the official views of the United States Air Force. The writer wishes gratefully to acknowledge considerable computational assistance from Mr. William L. Grafton.

Table 1

Rotated Factor Loadings for Stratified Sample of Technical School Students
Sample: 389 airmen assigned to technical schools
(Decimal points omitted throughout)

| Test Variables* | Factor I | II | III | IV | V | VI | $h^2$ |
|---|---|---|---|---|---|---|---|
| 1. Reading Vocabulary | 41 | 62 | 12 | 14 | 23 | 06 | 64 |
| 2. Arithmetic Computations | 72 | 15 | 17 | 04 | 22 | 05 | 62 |
| 3. Arithmetic Reasoning | 59 | 31 | 17 | 00 | 22 | 38 | 67 |
| 4. Pattern Analysis | 07 | 26 | −02 | 33 | 23 | 48 | 47 |
| 5. Mechanical Aptitude—2 | 14 | 18 | 00 | 27 | 64 | 16 | 56 |
| 6. Biographical Inventory No. 1 | 01 | −09 | 56 | 07 | −07 | −05 | 33 |
| 7. Memory for Landmarks | 25 | 25 | 16 | 39 | 12 | 32 | 42 |
| 8. Background for Current Affairs | 22 | 74 | 01 | 19 | 16 | 13 | 67 |
| 9. Biographical Inventory No. 2 | −01 | 15 | 53 | 22 | 10 | 16 | 39 |
| 10. Arithmetic Reasoning | 57 | 43 | 17 | −02 | 26 | 31 | 70 |
| 11. Aviation Information | 13 | 71 | −05 | 20 | 36 | 08 | 70 |
| 12. Dial and Table Reading | 49 | 32 | 19 | 38 | 19 | 25 | 62 |
| 13. Reading Comprehension | 29 | 69 | 05 | 10 | 28 | 23 | 70 |
| 14. Electrical Information | 06 | 53 | 04 | 18 | 59 | 19 | 70 |
| 15. Mechanical Principles | 02 | 10 | 02 | 07 | 38 | 47 | 38 |
| 16. Numerical Operations No. 1 | 70 | 04 | 17 | 36 | 08 | −18 | 69 |
| 17. Numerical Operations No. 2 | 78 | 11 | 07 | 28 | 08 | 03 | 71 |
| 18. General Mechanics | 01 | 42 | 08 | 00 | 63 | 14 | 60 |
| 19. Speed of Identification | 08 | 18 | 08 | 54 | 26 | 16 | 43 |
| | N | V | SE | P | ME | Vz | |

* Variables 1 through 4 are part scores of the AGCT. Variable 5 is also an AG test. Variables 6 through 19 are the Airman Classification Test Battery. A description of these tests is contained in Dailey (**2**, Appendix A).

comprehension (V); III. Socioeconomic background (SE); IV. Perceptual speed (P); V. Mechanical experience (ME); and VI. Visualization (Vz).

## Factor Loadings of Criteria

The validity coefficients of the 19 tests were available for 14 technical-school final-grade criteria. They were obtained from Dailey (**2**). Table 2 lists the criteria, N for each, and the means and standard deviations for the groups on certain representative tests (a relatively pure test of each of the six factors isolated). The standard deviations were examined in order to obtain an estimate of the relative variabilities of the groups on each factor. The factor content of the training criteria was estimated by Mosier's (**9**) extension of Dwyer's (**4**) method for obtaining the factor content of a variable not included in the original analysis. Table 3 gives the factor loadings and communalities of the criteria.[2]

The airmen were sent to the technical-training schools on the basis of their performance on the tests in the battery, and there is consequently considerable restriction of range in the tests, as well as on related tests for the selected groups. This restriction tends to reduce the validity coefficients and consequently the size of the factor loadings as compared with unselected groups. To estimate the size of the loadings for unrestricted samples, the values in Table 4 were calculated. The scores on the Airman Classification Battery are scaled to yield a standard deviation of 2.0 for unrestricted samples, and

[2] To reduce printing costs Tables 2 and 3 have been deposited with the American Documentation Institute. Order Document 3599 from American Document Institute, 1719 N Street, N.W., Washington 6, D. C., remitting $1.00 for microfilm (images 1 inch high on standard 35 mm. motion picture film) or $1.00 for photocopies (6×8 inches) readable without optical aid.

the A. G. tests are scaled to yield a standard deviation of 20.0 for similar groups. Using the standard deviations on the tests in Table 2 as estimates of the amount of restriction of each criterion group on the factors, the loadings for an unrestricted sample were estimated by the formula for correction for restriction-in-range, case 1 (**6**, p. 349), and are shown in Table 4. The loading of a criterion on a factor was considered the correlation between the criterion and the factor. An alternative and possibly more defensible procedure would have been to apply the corrections directly to the validity correlation coefficients.

It will be observed that the criteria have a considerable range of communalities. The very low communalities (e.g., for carpenter, radio operator ACS, and engineman-operator) are probably due largely to the unreliability of these criteria, although no direct evidence is available on this point.[3] It would be desirable to investigate the reliability of the grading systems used in these schools.

It should be kept in mind when evaluating the loadings that the criteria are grades in a course of study rather than performance on the job. The two highest loadings for airplane and engine mechanic (conventional) are on the mechanical-experience and visualization factors; for the jet mechanic on the mechanical-experience and verbal factors. The curriculum of the latter school might be examined to determine whether the stress on verbal material is necessary or even desirable.

The aircraft sheet-metal worker grades are loaded highest on visualization with a smaller loading on mechanical experience, whereas draftsman grades have their highest loading on mechanical experience with a somewhat lower loading on visualization. These loadings might also serve as a basis for evaluating these two curricula, as more visualization variance might well be expected in drafting grades and more mechanical-experience variance in grades in a sheet-metal course. The high numerical loading for the clerk-typist criterion is also

[3] If the reliabilities should prove to be satisfactorily high then, of course, the indication would be that there is considerable specific variance in the criteria and that they are not sufficiently covered by the factors in the battery.

Table 4

Estimated Factor Content of Technical Specialty Training Criteria Corrected for Restricted Variability
Sample: Graduates of the Technical Specialty Schools
(Decimal points omitted throughout)

| Criterion Group | Factor I | II | III | IV | V | VI | $h^2$ | N* |
|---|---|---|---|---|---|---|---|---|
| Weather Observer | 45 | 36 | 22 | 29 | 32 | 13 | 58 | 122–253 |
| Radio Operator, ACS | 22 | 29 | 23 | 24 | −02 | 16 | 27 | 108–114 |
| Radio Operator, General | 38 | 27 | 16 | 26 | 26 | 05 | 38 | 116–242 |
| A & E Mechanic (Conventional) | 31 | 26 | 04 | 15 | 74 | 44 | 93 | 245–384 |
| Aircraft Sheet-metal Worker | 28 | 35 | 07 | −01 | 39 | 53 | 64 | 136–243 |
| A & E Mechanic (Jet) | 25 | 42 | 21 | 06 | 58 | 31 | 72 | 254–300 |
| Carpenter | 17 | 12 | −01 | 13 | 18 | 21 | 14 | 190–362 |
| Control Tower Operator | 31 | 42 | 17 | 06 | 20 | 23 | 40 | 41–110 |
| Draftsman | 26 | 08 | 20 | 16 | 54 | 32 | 53 | 91–209 |
| Electrician | 57 | 38 | 17 | −03 | 36 | 09 | 64 | 74–164 |
| Engineman Operator | 16 | 08 | 12 | 06 | 43 | 16 | 26 | 69–144 |
| Clerk-Typist | 71 | 36 | 06 | 21 | −01 | 17 | 71 | 168–362 |
| Medical Corpsman | 30 | 57 | 05 | 15 | −07 | 15 | 47 | 426–834 |
| Radar Mechanic, General | 48 | 50 | 01 | 25 | 49 | 30 | 87 | 53–130 |
| | N | V | SE | P | ME | Vz | | |

* The sizes of the validation samples of the tests listed in Table 1 varied for each of the criterion groups within the limits shown.

somewhat surprising. A relatively higher verbal loading was expected.

## Discussion and Summary

The present study is intended to be exploratory and point out the possibilities in analyzing criteria as a method for better understanding the variance in technical-training course grades and other proficiency criteria. The steps in an ideal setup would be to:

1. Administer a comprehensive classification test battery to a large and representative sample.
2. Randomly assign members of the sample to courses of training so that the group taking each type of training would have an unrestricted range of talent.
3. Validate each test against a criterion of proficiency in the course of training. The establishment of reliable and valid criteria is in itself an important step.
4. Intercorrelate and factor-analyze the test battery.
5. Estimate the factor loadings of the criteria.

If desired, further steps might be to estimate the correlations among the criteria[4] and analyze the resulting correlations into clusters to form families of jobs requiring similar kinds and amounts of abilities.

With a comprehensive test battery and validities for large numbers of curricular and occupational criteria, it should be possible to put vocational guidance and occupational classification on a more objective basis.

*Received November 29, 1951.*

[4] By means of the formula for reproducing the correlation between two variables from their orthogonal factor loadings, $r_{12} = a_1a_2 + b_1b_2 + \cdots + n_1n_2$. Usually it is not possible to obtain the correlations between criteria directly, since a given group is rarely subjected to more than one course of training.

## References

1. Coombs, C. H., and Satter, G. A. A factorial approach to job families. *Psychometrika*, 1949, **14**, 33–42.
2. Dailey, J. T. Development of the Airman Classification Test Battery. *Air Training Command Research Bulletin 48–4*, November, 1948.
3. Dvorak, B. J. *Differential occupational ability patterns*. Minneapolis: Univ. of Minnesota Press, 1935.
4. Dwyer, P. S. The determination of the factor loadings of a given test from the known factor loadings of the other tests. *Psychometrika*, 1937, **2**, 173–178.
5. Guilford, J. P. Factor analysis in a test-development program. *Psychol. Rev.*, 1948, **55**, 79–94.
6. Guilford, J. P. *Fundamental statistics in psychology and education*. (Second edition.) New York: McGraw-Hill, 1950.
7. Gulliksen, H. Intrinsic validity. *Amer. Psychologist*, 1950, **5**, 511–517.
8. Lawshe, C. H., Dudek, E. E., and Wilson, R. F. Studies of evaluation. VII. A factor analysis of two point rating methods of job evaluation. *J. appl. Psychol.*, 1948, **32**, 118–129.
9. Mosier, C. I. A note on Dwyer: The determination of the factor loadings of a given test. *Psychometrika*, 1938, **3**, 297–299.
10. Shartle, C. L. *Occupational information*. New York: Prentice-Hall, 1946.
11. 3309th Research and Development Squadron. *Research report No. 10*. San Antonio, Texas, Lackland Air Force Base, January 1949.
12. Viteles, M. S. *Industrial psychology*. New York: W. Norton, 1932.
13. Zerga, J. E. Job analysis, a resumé and bibliography. *J. appl. Psychol.*, 1943, **27**, 249–267.

# A Psychological Study of Occupational Adjustment *

**Alastair Heron**

*Medical Research Council, London, England*

This paper reports a study of male unskilled factory workers. The aim may be stated as follows: by means mainly of objective psychological tests to study the relationships which may thereby be shown to exist between various aspects of the personalities of a group of unskilled male factory workers, and the extent to which they appear to be meeting the demands of the job situation.

## Population Studied

The population consisted of an intact section of a basic production department in a medium-sized factory. The 80 men concerned, varying in age from 22 to 64 years, were unskilled operatives on individual piecework who had been in that department for periods ranging from 1 to 30 years, the median falling at 10 years. As raw material accounted for nearly 90 per cent of the cost of the product, the firm offered attractive wage and bonus rates. The job is one which over the years would select men who were in the main physically fit steady workers unlikely to be exceptional in many respects. It is the kind of job in which Russell Fraser (**2**) found the lowest incidence of neurosis.

Study of the task—pouring molten lead into hand-operated moulds—suggests that only minimal intellectual equipment would be required, and that some obsessionals might adapt well. Beyond that it was not meaningful to hypothesize in terms related to the demands of the job, except insofar as it is generally believed that emotional instability is reflected in poor job adjustment.

## Personality Variables

A battery of 22 individually administered objective tests was used, designed to cover such aspects of personality as general mental ability, emotional stability, temperament, and dexterity. Full details of the tests, the results obtained and the methods of statistical analysis will be found elsewhere (**3**). For the present purpose it will suffice to say that, with age partialled out, the matrix of product moment correlations between the tests was factor-analyzed, using Burt's simple summation method. The analysis yielded four significant factors accounting for 30 per cent of the variance. After five orthogonal rotations, three of these factors were readily identified in the light of previous data about the tests concerned.

Factor scores were obtained by simple addition of the individual's normalized percentile scores on the four tests having the highest loadings on the factor in question, in the final rotated solution. In this way *no weight has been given to the actual loadings themselves*. The factor analysis has in effect been used to select small groups of tests which have been shown to be related to one another in a psychologically meaningful way. Its arithmetic is then abandoned, and relationships between the "factor scores" and the criteria are expressed as zero-order coefficients.

## Criteria of Occupational Adjustment

Details of the establishment of these criteria will be found elsewhere (**4**). In view of their decisive importance, however, a description of each is considered necessary here.

*Productivity.* This was an index based on the individual's average productivity over a period of 67 weeks, as obtained from shop-floor records. It appeared to be so closely related to actual output (checked for 45 men over 26 weeks on a particular type of casting) as to justify considerable confidence in its use. The principal criticism to which it should be subjected is that it may suffer from artificial limitation of range due to external variables. It is free from the familiar criticism in terms of the unknown effect upon production which

* The author is a member of the Unit for Research in Occupational Adaptation, which is based in the Maudsley Hospital, Denmark Hill, London, S.E. 5. This paper was presented in condensed form to the 10th International Congress of Psychotechnics at Gothenburg, Sweden, in July 1951. It is a revision of part of a Ph.D. thesis accepted by the University of London in 1951.

can result from direct observation of the worker or the installation of special counting devices. As these men were working on an individual piecework basis, such an index seems to provide a measure of occupational effectiveness in terms of the individual's "ability" to earn his living relative to that of others doing the same work. This does not, however, take into account the incentive differences arising from social, familial and economic pressures. Two men might easily possess the same hypothetical ability to carry out the working-operation at a high average rate, but one might show an actual rate much lower than the other because, for example, his wife was earning and he had no children to support. All but one of the men in the present study were married and a considerable proportion had children under school-leaving age. The latter variable was found to correlate with Productivity +.226 (P = .05), thus illustrating the point under discussion. It seems obvious that no collection of psychological assessments could hope to account among them for more than a small proportion of the variance of this criterion.

*Job Adjustment.* For this the author obtained a rating on a normally distributed 5-point scale, combined from the independent ratings of six supervisors on each of two occasions five weeks apart. The average consistency coefficient of the raters was .78, and careful statistical analysis established the unidimensionality of the criterion. It may be defined as a measure of *the extent to which a man is a source of concern to his supervisors.*

This criterion differs from a "merit rating" as generally used. The latter is most usually a means whereby men are considered for promotion in status, responsibility or earnings, and the rater's judgments are deliberately oriented in such a way as to maximize the predictive power of the rating for these purposes. In the present situation, no such merit rating system was in existence; the procedure used was to obtain an assessment of the extent to which each man was a source of concern to the supervisor. Further, the role of the investigator as a member of a research unit was already well established, and his independence of management recognized to a very considerable extent. The influence of "halo" resulting from familiarity was not, however, wholly absent; this could not readily be avoided when studying a well-established group whose experience in that department ranged from 1 to 30 years. In general, the criterion may perhaps be regarded with some justification as being satisfactory for the purpose.

## Results

In view of the significant correlation with *age* of various tests in each factor score, its effects have been held constant by partial correlation when computing the relationship of each factor score with the Productivity criterion. When the Job Adjustment criterion is involved, both *age* and *experience* in the shop are held constant, as the latter variable correlates +.4 with this criterion.

For the sake of clarity the results are presented in summary form, and only those coefficients are shown which are significant at the .05 level or better. Short titles of tests are provided in order to facilitate reference to details available elsewhere (**3**).

1. Factor I. "General Mental Ability" (Dominoes Non-verbal + Letter Series + Paper Formboard + Vocabulary) showed no significant relationship with either criterion.

2. Factor II. This factor is not named, but is suggestive of "Hysteric Tendency" or "Neurotic Extraversion" (low Hand Persistence + low Leg Persistence + many Food Aversions + high "neurotic" score on Word Connection List) and also showed no significant relationship with either criterion.

3. Factor III. "Emotional Instability" (many Worries + much Static Ataxia + many Annoyances + many Interests) correlated with Poor Job Adjustment +.45 (P = .001).

4. Factor IV. "Speed of Approach" (Finger Dexterity + Quick Approach to Time Test + Manual Dexterity + Speed on Track Tracer) showed the following relationships: (a) with Productivity, +.25 (P = .05); and (b) with Good Job Adjustment, +.28 (P = .01).

5. An *unweighted* combination of Factor III (reversed) and Factor IV, using the "pooling square" technique, gave a correlation with Good Job Adjustment of +.53.

## Discussion

It may be concluded formally that the null hypothesis was confirmed in respect of general mental ability, as assessed by non-verbal, verbal-educational and visuo-spatial tests. In view of the low level of intellectual demand implicit in the actual job, and the number of men employed on it who were shown to be well above average in general mental ability, this finding is of some importance. If it were true in this instance, as suggested by Wyatt and Langdon (**6**), that such men would be handicapped by their excess of unneeded intellectual potential, this should have been reflected in a small negative relationship with one or more criteria of adjustment; no such relationships were found. It may be, of course, that these men exhibit adaptation to other aspects of the total situation which favor them in a way which they had not encountered on jobs apparently more in line with their intellectual potentialities. Conversely, it also seems of some importance that men of extremely low mental ability—equivalent in several cases to that of high-grade mental defectives—appear able to do this job satisfactorily. This supports the conclusion reached by Tizard and O'Connor (**5**), when considering the employability of high-grade mental defectives, that monotonous work of a relatively simple kind may suit them well.

If the hypothesis be advanced that "emotional instability" as evidenced by objective tests is related systematically to "poor job adjustment" as rated by supervisors, it is sustained by the findings. It is not sustained in respect of productivity. The four tests having the highest loading on this factor are all associated with what has been described as the "dysthymic" rather than the "hysteric" end of the temperament continuum in neurotics (**1**). This includes the anxiety states, obsessional tendency and reactive depression. It seems that in the sample studied it is this heterogeneous group of relatively unstable men who tend to be a source of concern to their supervisors. Their handicap is not, however, sufficient to affect their productive capacity over a long period.

No hypothesis is required to provide a basis for the unsurprising finding that a tendency to "Speed of Approach to a Task," as measured by the Factor IV score, is related to Productivity. If, as has been suggested, this may be regarded as a characteristic mode of personality, then for a man engaged on individual piecework its deficiency appears to constitute a handicap. At this point we may recollect the attenuating effects on such a relationship of the defects of the criterion; it may well be that a coefficient of .25 represents in this instance a more serious handicap than would appear from its size.

The combination of a high tendency to "emotional instability" of an anxious or depressive type, and a slow characteristic approach to a task, is found to be related to supervisory assessment of job adjustment at a level which, without the aid of weights derived from multiple correlation, is high enough (.53) to be regarded as evidencing a genuine handicap in the situation studied.

## Summary

1. This paper reports the relationships found between personality variables and occupational adjustment in a sample of 80 male unskilled factory workers.

2. Personality variables were derived from a factor analysis of the intercorrelations between 22 individually-administered objective tests.

3. Two occupational criteria were specially prepared for the investigation; one was a measure of productivity, the other of the extent to which men were a source of concern to their supervisors.

4. Some significant relationships were found and are discussed briefly from various points of view.

*Received January 18, 1952.*

## References

1. Eysenck, H. J. *Dimensions of personality*. London: Kegan Paul, 1947.
2. Fraser, R. *The incidence of neurosis among factory workers*. Rep. No. 90, Ind. Hlth. Res. Bd. London: H. M. Stationery Office, 1947.
3. Heron, A. The objective assessment of personality among factory workers. In press.
4. Heron, A. The establishment for research purposes of two criteria of occupational adjustment. *Occup. Psychol.*, 1952, **26**, 2, 78–85.
5. Tizard, J., and O'Connor, N. The employability of high-grade mental defectives, II. *Amer. J. ment. Def.*, 1950, **55**, 144–157.
6. Wyatt, S., and Langdon, J. N. *Fatigue and boredom in repetitive work*. Rep. No. 77, Ind. Hlth. Res. Bd. London: H. M. Stationery Office, 1937.

# *How Supervise?* Scores Before and After Courses in Psychology

**Frederic R. Wickert**

*Michigan State College*

*How Supervise?* is a device developed by File and Remmers for the measurement of attitudes and understandings necessary for supervisory success. Karn (**2**), in summarizing the rather inconclusive evidence for the validity of this questionnaire, states, ". . . Further studies appear to be in order before the question concerning the universal validity of the test can be settled."

Karn then reported an investigation of his own which he felt had a bearing on the validity of *How Supervise?*. He administered Forms A plus B of this test to 108 college students in their junior year, before and after a course in psychology. He also administered the two forms of this test to a comparable control group of 104 students before and after a course in English literature. Results showed that the two groups achieved the same mean scores on the test at the start of the psychology and English courses, respectively. When the test was readministered at the end of each course, however, the psychology (experimental) group had gained significantly, but the English (control) group had not. Karn further reports that Form B alone (as well as Forms A plus B) was sufficiently sensitive to show a significant difference for the psychology (experimental) group on initial and final testings.

The present study was designed, from one point of view, to supplement the Karn study by providing further evidence on the validity of *How Supervise?*. From another point of view, if one assumes that *How Supervise?* measures the attitudes and understandings that enter into good human relations, the present study was concerned with discovering the degree to which such attitudes and understandings had been taught in various types of psychology courses at the college level. From the first point of view, then, the present study was a test validation project; from the second point of view it was an educational research project. The reader may take his choice.

## Procedure

Four groups of students at Michigan State College served as subjects.[1] Group 1 was made up of 68 students in the begining course in general psychology. The majority of these students were sophomores with a sprinkling of freshmen as well as a few juniors and seniors. They came from many departments of the College. The content of this one-quarter course was that traditionally covered in such courses: perception, emotion, learning, motivation, etc. No attempt was made to discuss psychological factors in human relations problems.

The 77 students in Group 2 were also primarily sophomores. No freshmen were included, but there was a somewhat larger proportion of juniors and seniors than in the general psychology course. This second course was the beginning course in industrial psychology, and, as might have been expected, about half the students came from the School of Business and Public Service and the other half from the School of Science and Arts. The course in general psychology was a prerequisite for this course, so it represents the second stage in a hierarchy of courses. About one-fifth of the course was directly devoted to teaching the psychology of human relations in industry; the remainder of the course was concerned with such traditional topics in industrial psychology as selection testing, merit rating, training, job analysis and job evaluation.

The 51 students in Group 3 were about evenly divided between juniors and seniors. They were taking a course completely devoted to the study of psychological factors in human relations in industry. The prerequisite for this course was the beginning course in industrial psychology. Hence it was the third course in the hierarchy.

[1] Group 1 was 41% women, Group 2, 17%, Group 3, 15%, and Group 4, 29%.

For Group 4, 31 students taking either or both of two advanced courses in industrial psychology were used. These students were all seniors. These two courses, which had the course in the psychology of human relations in industry as one of their prerequisites, were concerned with: (1) how to do various kinds of interviewing in industry and government situations; and (2) training and supervising in industry and government. The courses depended on and referred to psychological factors in human relations but did not specifically emphasize them or elaborate upon them.

These four groups, then, represented each of four stages of a hierarchy of successively more advanced courses in psychology, the first of which had little or no human relations training content, the second with some, the third almost completely, and the fourth relatively little except as an assumed foundation. The problem of this study was to work out the relationship between training in human relations and performance on *How Supervise?*. Fortunately, Karn, in the study already referred to, had demonstrated that a control group, under circumstances roughly comparable to those of the present study, showed no appreciable gain between initial and terminal testing on *How Supervise?*. Accordingly, no series of control groups matched with each of the four experimental groups were utilized in the present study. It was assumed, perhaps not with complete justification, that such control groups would have shown no gains just as Karn's control group did not.

The details of administration of the test are as follows. In each of the four groups during the first week of the same college quarter, Form A was given to one half the group and Form B to the other half. During the last week of the quarter, Form B was given to those who were initially tested on Form A, and Form A was given to those initially tested on Form B. This reversing of forms was done since File said, "Forms A and B are closely equivalent." He also added, ". . . Norms of 'One Form Only' may be applied to either Form A or Form B." Furthermore he suggested using the different forms if "repeated testing is anticipated" (**1**).

## Results

Since Form A was given first to half of each group and was followed by Form B, and the opposite procedure was followed for the other half of each group, it was possible to check the comparability of the two forms in detecting the effects of training. Table 1 shows this comparison. It will be seen that the means of both Forms A and B, when given first, were practically identical. It will also be seen that when Form A was followed by Form B, a significant gain occurred. However, when Form B was followed by Form A, some gain, but not a statistically significant one, resulted. This finding suggested that the two forms, though identical on first administration, were not equally sensitive to picking up the effects of the kind of human relations training involved in this study. Notwithstanding this finding, it was decided to include in the analysis of results all the cases, that is, to include those who took Form B first followed by Form A as well as those who took Form A first followed by Form B.

The principal results are shown in Table 2. This table compares initial and terminal scores on *How Supervise?* for each of four successively more highly trained groups. The table indicates that there was some gain in all four groups. Group 3, which received the most

Table 1

Comparison between Forms A and B of *How Supervise?* When Used in Initial and Terminal Testing, All Four Groups Combined

| | Form A Given First and Form B Second | Form B Given First and Form A Second |
|---|---|---|
| N | 114 | 113 |
| Mean of form given first | 48.07 (Form A) | 48.11 (Form B) |
| S.D. | 9.62 | 8.88 |
| Mean of form given second | 53.00 (Form B) | 49.16 (Form A) |
| S.D. | 7.58 | 7.90 |
| Difference between means | 4.93 (Gain) | 1.05 (Gain) |
| "t" | 4.98* | 1.01 |

* Significant at 1% level. (A "t" of 2.6 is needed for significance at 1% level and 1.99 at 5% level.)

Table 2

Comparison of Initial and Terminal Scores on *How Supervise?* for Four Successively More Highly Trained Groups

| | Group | N | | Mean | S.D. | Before vs. After "t" | After vs. Next Before "t" |
|---|---|---|---|---|---|---|---|
| 1. | General Psychology | 68 | Before | 44.74 | 9.58 | 1.01 (gain) | |
| 2. | | | After | 46.26 | 7.65 | | .03 (loss) |
| 3. | Sophomore Industrial Psychology | 77 | Before | 46.22 | 9.10 | 2.44* (gain) | |
| 4. | | | After | 49.48 | 7.30 | | 2.31* (gain) |
| 5. | Psychology of Human Relations in Industry | 51 | Before | 52.39 | 6.66 | 3.58** (gain) | |
| 6. | | | After | 56.67 | 5.21 | | 2.32* (loss) |
| 7 | Senior Industrial Psychology | 31 | Before | 53.00 | 7.74 | 1.81 (gain) | |
| 8. | | | After | 56.13 | 5.44 | | |

* Significant at 5% level.
** Significant at 1% level.

direct training in human relations principles, showed the greatest gain (significant at the 1% level). Group 2, which received some human relations training, showed the next greatest gain (significant at the 5% level). Group 4, which received additional human relations training indirectly only, showed considerable gain but the N was too small to show statistical significance. Group 1, the training for which included only occasional and then only superficial references to human relations principles, showed the least gain (not statistically significant).[2]

The results in Table 2 also suggest the operation of certain factors other than training in human relations. At the time this project was carried out, there were, during each of the 3 quarters of the regular academic year, about 20–22 sections of general psychology of about 80 students each, 5–6 sections of sophomore industrial psychology of about 80 students each, and 2 sections of the psychology of human relations in industry of from 25 to 30 students each.[3] Two of the three quarters, there were generally 2 different classes in senior industrial psychology of 20 students each. This means that only perhaps a fourth of the students in general psychology went on to sophomore industrial psychology. For many students, these two courses were required. Group 3, however, was more highly self-selected; the course in the psychology of human relations was not required so that it was usually limited to highly interested students. Generally, students took the courses represented by Groups 1, 2, and 3 closely following one another in time. There was usually a considerable lapse of time, often 3 to 5 quarters, before taking senior industrial psychology.

With the above administrative factors in mind, it is not difficult to interpret certain of the results in Table 2. Students in sophomore industrial psychology entered that course at just about the level of training in human relations they had attained in their general psychology course. They did not appear to be

[2] An explanation of the appreciable gain in the test scores of Karn's general psychology group before and after training and the negligible gain of the general psychology group at Michigan State College deserves to be mentioned. Karn states regarding the Carnegie Institute of Technology course, "the emphasis . . . was . . . upon the application of psychological principles to adjustment problems, particularly those likely to be encountered in human relations situations in industry," while the content of the course at Michigan State College, as was mentioned earlier in this paper, was more orthodox.

[3] It should be pointed out that the one section of general psychology and one section of sophomore industrial psychology were assumed to be representative of the total group. Groups 3 and 4, however, included all the students in the courses covered in this study.

very much different in their knowledge of human relations principles from the mass of students finishing the one beginning course in general psychology. The relatively small group of students who went on to the psychology of human relations in industry, however, started significantly above the mean (as shown by the entry in the column headed "after vs. next before 't' "). The fact that the still smaller group (Group 4) in the senior industrial psychology courses started and ended at about the same level as Group 3 is not easy to interpret. It is possible that generally the poorer students from the psychology of human relations classes went on into the senior industrial psychology. It is also possible that these seniors might have forgotten some information in the 3 to 5-quarter time gap between their having taken the psychology of human relations course and the senior industrial psychology course. A further possibility is that these students were scoring so close to the upper possible scores (about the 95th percentile on Files' "lower level supervisor" norms and about the 70th percentile on the "higher level supervisor" norms) that it was difficult for a group average to stay consistently high.

## Discussion

The results of this study showed first that the two forms of the File-Remmers questionnaire might well be equivalent *before* a specific course of training in human relations principles but were not equivalent in their sensitivity to changes brought about by such a course of training. One form (Form B) proved to be quite sensitive but the other form (Form A) was not. Some further analyses of the kinds of items in the two forms, of the content of the training courses, and of the data in this project (not reported here) failed to disclose any very good reasons for this finding.

Notwithstanding the differential sensitivity of the two forms, the File-Remmers test was found to have considerable validity in the sense that gains in the scores before and after a course of training corresponded closely to the known content of the training courses. The greater the human relations content, the greater was the gain.

If on the other hand the File-Remmers test is assumed to have validity, it could then be stated that the training was having a significant effect in inculcating the attitudes and understandings underlying good human relations.

This study, then, supplements that of Karn. In general, it shows successive gains in knowledge of human relations principles through a series of successively more advanced courses in psychology. It also, to use Karn's words, is "a contribution towards the establishment of the universal validity of the test since it deals . . . with college students under academic instruction. Previous studies, which have shown positive gains with training, have been concerned with the effects of specific industrial training programs among on-the-job supervisory personnel. The demonstration of improved scores under a variety of training conditions indicates that the responses are a result of the application of basic principles to the problems rather then the acquisition of specific answers to the questionnaire items."

## Summary

Four groups of college students, each group at a successively more advanced level of training with respect to knowledge of human relations principles, were given the File-Remmers test, *How Supervise?*, before and after each of four successively more advanced courses in psychology. Control groups were not used since an earlier study by Karn, under roughly comparable circumstances, had already shown that while an experimental group of college students gained in knowledge of human relations principles, a control group did not gain appreciably. It was then assumed, perhaps not with complete justification, that if control groups matched for each experimental group in the present study had been utilized, they would not have shown the gains made by the experimental groups.

Results showed that the two forms of the test were closely equivalent before training, but Form A following Form B was considerably less sensitive in detecting the effects of human relations training than Form B following Form A.

Results also showed that gains in mean scores for the four groups corresponded closely to the amount of human relations training believed to have been included in each of the four courses. Moreover, in general the mean score level of each successively more advanced group at the beginning of training corresponded closely to the level of human relations knowledge that each group could have been expected to have by virtue of the degree of human relations training it had received in the past. These results may be interpreted in either one, but not both, of two ways. If it is accepted that the courses in psychology were inculcating the attitudes and understandings that enter into good human relations, then the results may be taken to show some validity for *How Supervise?* as a test of such attitudes and understandings. On the other hand, if *How Supervise?* is accepted as a test of the attitudes and understandings that enter into good human relations, then the results may be taken to show that the psychology courses, particularly the ones designed to do so, were in fact successfully inculcating such attitudes and understandings.

*Received January 8, 1952.*

## References

1. File, Q. W., and Remmers, H. H. *How Supervise?* (revised manual). New York: 1948, Psychological Corporation, pp. 8.
2. Karn, H. W. Performance on the File-Remmers Test, *How Supervise?*, before and after a course in psychology. *J. appl. Psychol.*, 1949, **33**, 534–539.

# Measurement of Supervisory Ability

**Gerald C. Carter**

*University of Illinois*

A battery of tests which consisted of *The Personal Audit; How Supervise? Form A and B; Mechanical Comprehension Test Form AA; Kuder Preference Record;* and the *Otis Quick Scoring Mental Ability Test* was given to the foremen and assistant foremen in two metal fabricating plants. Ratings by fellow supervisors were used as criteria of supervisory ability. Group A was composed of 15 foremen in the larger of the two plants. Group B was composed of 16 assistant foremen in the same plant and Group C was composed of 17 foremen in the smaller of the two plants. Some of the supervisors did not take all of the tests. Therefore, the N varied somewhat from test to test.

## Criterion

The number of fellow supervisors who did the rating in the larger plant was 27. The number of fellow supervisors who did the rating in the smaller plant was 15.

The ratings were made by an open middle forced distribution technique. Each rater was given a sheet with the following directions: "The results of this rating will be kept confidential. You are a plant superintendent given the task of selecting five supervisors from the following list of names. Write the names of the five men you would choose first. Write the names of the five men you would consider last. Omit your own name. Try not to let personal feelings affect your decision. Do not sign this paper."

The names of the supervisors in the three groups were placed on separate sheets. A score for each supervisor was obtained by assigning three points for each time he was mentioned in the first category, 1 point for each time he was mentioned in the last category, and 2 points for each time he was automatically placed in the open middle. The total of these points was then divided by the number of raters for that particular group and multiplied by 10 to remove the decimal point.

The uncorrected split-half reliabilities of the ratings were computed and found to range from .75 for group B to .82 for group A. The reliability was .85 for the three groups combined. The criterion is more reliable than is usually found because it is a "pooled rating."

## Results

The coefficients of correlation between the criterion and scores on the *Mechanical Comprehension Test, Otis Quick Scoring Mental Ability Test,* scores on each part of *The Personal Audit,* scores on each part of the *Kuder Preference Record* and scores on each part and total scores for Form A and B separately and combined for *How Supervise?* are shown in Table 1.

The coefficients of correlation shown in Table 1 included all supervisors who had taken each test. Therefore, these coefficients differ somewhat from those shown in Table 2 because these were recomputed to include only those supervisors who had taken all of the three best predictors for each group. If a supervisor had failed to take one of the tests used to obtain the multiple correlation for a group, his score had to be removed from correlation of the other two tests so that identical sets of individuals were used for each of these three tests.

In order to determine the extent to which the three best predictors measured supervisory success as indicated by the ratings, coefficients of multiple correlation and their standard errors were computed for each group and for the three groups combined. Regression coefficients and the regression equation were established for the three groups combined.

The regression coefficients show relative weights of each of these three measures in contributing to the supervisory ability as measured by the pooled ratings of fellow supervisors. The regression equation may be used to predict the rating of any particular supervisor from the knowledge of his scores on each of the three best predictors. For in-

stance, if a prediction of the rating is to be made for Mr. X, his score of 21 on part three of *How Supervise? Form A*, his score of 31 on the artistic section of the *Kuder Preference Record*, his score of 32 on the *Mechanical Comprehension Test* are multiplied by the relative weights of these three factors. The sum of these products added to the constant gives the predicted rating. The equation for the predicted rating (P) for Mr. X is:

$$P = .244 \times 21 + .023 \times 31 + .233 \times 32 + 5.107 = 18.4.$$

It can be predicted that Mr. X should have a rating of 18.4. An examination of the actual rating indicates that he received a pooled rating of 17 by his fellow supervisors. The coefficient of multiple correlation indicates the degree of accuracy of the prediction of the rating from the regression equation. These coefficients were .74 for group A, .80 for group B, .80 for group C and .84 for the three groups combined.

The coefficient of multiple correlation for the three groups combined (.84) being quite high might give the erroneous impression that the predictors used to obtain it are highly effective for each of the three groups. However, this is not true. Two of these predictors are ineffective for group A.

Table 1

Coefficients of Correlation between Criterion and Test Scores

| | | Group A | | Group B | | Group C | | Total Group | |
|---|---|---|---|---|---|---|---|---|---|
| | | N | r | N | r | N | r | N | r |
| Personal Audit | | | | | | | | | |
| Seriousness | | 13 | −.63 | 15 | −.16 | 14 | .09 | 42 | −.20 |
| Firmness | | 13 | .10 | 15 | .13 | 15 | −.03 | 43 | .06 |
| Tranquillity | | 13 | −.42 | 15 | −.09 | 15 | −.03 | 43 | −.16 |
| Frankness | | 13 | .12 | 15 | .10 | 15 | −.01 | 43 | .05 |
| Stability | | 13 | −.06 | 15 | −.09 | 15 | −.05 | 43 | −.06 |
| Tolerance | | 12 | .13 | 15 | .52 | 14 | .03 | 41 | .18 |
| Kuder Preference Record | | | | | | | | | |
| Mechanical | | 14 | .04 | 15 | .04 | 15 | .15 | 44 | .11 |
| Computational | | 14 | −.07 | 15 | .03 | 15 | −.24 | 44 | −.13 |
| Scientific | | 14 | .13 | 15 | .09 | 15 | .21 | 44 | .15 |
| Persuasive | | 14 | −.40 | 15 | .04 | 15 | −.36 | 44 | −.24 |
| Artistic | | 14 | .64 | 15 | .38 | 15 | .12 | 44 | .32 |
| Literary | | 14 | −.05 | 15 | −.28 | 15 | .44 | 44 | .16 |
| Musical | | 14 | .04 | 15 | −.10 | 15 | −.44 | 44 | −.16 |
| Social Service | | 14 | −.28 | 15 | −.29 | 15 | .15 | 44 | −.09 |
| Clerical | | 14 | −.16 | 15 | .44 | 15 | −.47 | 44 | −.11 |
| *How Supervise?* | | | | | | | | | |
| Form A | Part I | 13 | −.19 | 13 | .16 | 17 | .15 | 43 | .02 |
| | Part II | 13 | .08 | 13 | −.06 | 17 | −.06 | 43 | −.03 |
| | Part III | 13 | −.04 | 13 | .46 | 17 | .63 | 43 | .39 |
| | Total | 13 | −.06 | 13 | .27 | 17 | .51 | 43 | .22 |
| Form B | Part I | 13 | −.24 | 15 | .03 | 17 | .48 | 45 | .14 |
| | Part II | 13 | −.02 | 15 | .61 | 17 | .43 | 45 | .30 |
| | Part III | 13 | −.04 | 15 | .46 | 17 | .35 | 45 | .39 |
| | Total | 13 | −.02 | 15 | .53 | 17 | .55 | 45 | .31 |
| Form A & B | Part I | 11 | −.17 | 12 | .23 | 17 | .39 | 40 | .11 |
| | Part II | 11 | .05 | 12 | .46 | 17 | .27 | 40 | .20 |
| | Part III | 11 | .10 | 12 | .66 | 17 | .59 | 40 | .34 |
| | Total | 11 | .03 | 12 | .57 | 17 | .58 | 40 | .27 |
| Otis Test | | 14 | .20 | 15 | .00 | 16 | .19 | 45 | .10 |
| Mech. Compr., AA | | 13 | −.24 | 16 | .15 | 15 | .39 | 44 | .29 |

Table 2

| | Uncorrected Split-Half Reliability of Criterion | Three Best Predictors with Coefficients of Correlation with Criterion | Number of Supervisors Taking These Three Tests | Coefficient of Multiple Correlation | Standard Error of Multiple Correlation |
|---|---|---|---|---|---|
| Group A | .82 | Kuder Preference Record, Artistic ($r = .66$)<br>Personal Audit, Seriousness ($r = -.58$)<br>Personal Audit, Tranquillity ($r = -.40$) | 12 | .74 | .13 |
| Group B | .75 | *How Supervise?*, Part III, Form A+Form B ($r = .66$)<br>Personal Audit, Tolerance ($r = .58$)<br>Kuder Preference Record, Social Service ($r = -.37$) | 12 | .80 | .10 |
| Group C | .80 | *How Supervise?*, Part III, Form A ($r = .65$)<br>Kuder Preference Record, Clerical ($r = -.47$)<br>Kuder Preference Record, Musical ($r = -.44$) | 15 | .80 | .09 |
| Total | .85 | *How Supervise?*, Part III, Form A ($r = .63$)<br>Kuder Preference Record, Artistic ($r = .13$)<br>Bennett Mechanical Comprehension ($r = .62$) | 36 | .84 | .05 |

Although all of the coefficients of multiple correlations were surprisingly high, being in the order of the magnitude of the reliabilities of the criteria and of the tests, this should not be construed as evidence that any of the sets of three best predictors would necessarily be effective for predicting supervisory ability in general. The effectiveness of the predictors varied greatly from group to group. This is very probably due to the unstable zero order correlations and the tendency of multiple correlation to unduly magnify chance errors of measurement. Because of this variation a cross validation study should be made for every supervisory group for which predictors are desired.

The results presented indicate that, at least in some instances, supervisory ability can be measured fairly accurately by psychological tests by selecting the best predictors from a relatively large number of likely indices (29 in this study). It is also necessary to have reliable criteria such as pooled ratings by a large number of raters who know the supervisors being rated exceedingly well.

*Received February 13, 1952.*

# Studies of Group Behavior: Factors Associated with the Productivity of Groups [1]

**John G. Darley**

*University of Minnesota*

**Neal Gross** [2]

*Harvard University*

**and**

**William C. Martin** [2]

*University of Illinois*

This is one of a series of papers reporting the results of a study of the relations among selected sociological and psychological variables in the behavior of small, organized groups. The present phase of the total study deals with the analysis of factors related to group productivity. The groups studied comprised thirteen women's small residence units organized as a cooperative housing project under the auspices of the University of Minnesota. Background data on these small groups appear in items (**1**) and (**2**) of the References.

## The Criterion of Productivity

During the course of the investigation, an attempt was made to get an estimate of the relative productivity of the groups when engaged in a common and meaningful task as a criterion measure against which to correlate certain group variables.

While the cooperative housing project served primarily to provide room and board for the students at a lower cost than was elsewhere available to them, the University sought, by giving it full recognition as a student activity, to make the residence experience a worth-while phase of personal growth and development for them. Problems of cooperative living were stressed in their broader implications for society generally. Against this background of a consciously-shared experience, the experimenters proposed that each house prepare a "plan for better cooperative living in the Village." Each house was allowed one month during the spring term for the preparation of its report, as a house project and in conformity to a suggested standard outline. Cash prizes of $100, $75, $50, $35, and $25 were announced for the five top-ranking houses in the contest. Faculty members who agreed to serve as judges were asked to make an over-all ranking of the thirteen reports in the form of a single total judgment, and the criteria to be used by the judges were also given to the students at the start of the contest.

## The Independent Variables

The predictor variables in such a study must bear some explicit or implicit relation to tentative hypotheses regarding factors associated with group productivity. As a first approximation, the problem was formulated in the following manner:

All other things being equal, the productivity of a group will depend upon the extent to which: A. the group goal is accepted; B. the previous group experience has been successful and satisfying; C. appropriate use is made of differential skills within the group membership; D. leadership is both accepted and persistent; and E. there are available

[1] The studies to be reported in this and subsequent articles, under the general title of *Studies of Group Behavior*, have been carried on by the Laboratory for Research in Social Relations at the University of Minnesota under grants from the Graduate School; the College of Science, Literature, and the Arts; and from the Carnegie Corporation. They are part of a larger research program dealing with the broad problems of social responsibility.

[2] Members of the Minnesota Laboratory at the time of the research.

within the group individuals of requisite ability or special skill.

Before describing the independent variables assumed to measure the behavior needed to test the above hypotheses, it is necessary to describe briefly one additional source of data, the participant-observer reports. The University-assigned counselor in each house was asked to play the role of participant-observer during the month in which the reports were prepared. The tasks of the participant-observer were as follows: to prepare a narrative account of each house meeting held concerning the contest; to report on the attendance at each such meeting; to rate every participant at these meetings on a five-step scale for specified behavior.

The narrative accounts of the meetings were designed to cover: enthusiasm for the task to be accomplished by the group; strength and sources of leadership in the task; type of leadership (democratic vs. authoritarian); evidences of conflict within the group; and organizational efficiency with which the group carried out the task. The participant-observers rated each student present at each meeting on a five-step scale for: extent of participation in the meeting; usefulness of suggestions; cooperation; willingness to assume responsibility; and leadership.

The narrative reports submitted by the participant-observers were later independently analyzed by two raters, who used simple five-step scales (with step 1 being arbitrarily assigned as the "good" score) to assess: strength of leadership; type of leadership ("autocratic vs. democratic"); efficiency of organization; evidences of interpersonal conflict; enthusiasm; and amount of leadership provided by the counselor (who was also the participant-observer). These assessments, based on all narrative reports for each house, represented composite analyses of house behavior during the entire month of the contest; final house ranks were assigned by averaging the ranks assigned by the two raters for a particular variable for each meeting held by the group during the month of the contest.

Returning now to the theoretical formulation of the problem, it is in order to list the kinds of data derived as measures of the independent variables related to productivity. These are listed herewith, under the five general behavior categories:[3]

A. Acceptance of group goal:
   1. Volume of participation (attendance records)
   2. Extent of participation (as rated by participant-observers)
   3. Assumption of responsibility (as rated by participant-observers)
   4. Enthusiasm for the task (as judged by case readers)

B. Success or satisfaction with previous group experiences:
   5. Satisfaction with Village life (from autobiographical reports)
   6. House dislike ratio (from end-of-year sociometric schedules)
   7. House ratio of in-group to out-group choices for "social support" (from start-of-year sociometric schedules)
   8. House ratio of in-group to out-group choices for "confidantes" (from start-of-year sociometric schedules)
   9. House ratio of in-group to out-group choices for "intimate friends" (from end-of-year sociometric schedules)
   10. Evidence of personality conflict (as judged by case readers)

C and E. Availability of essential skills and appropriate use of such skills:
(These two dimensions are treated together in the following list.)
   11. House scholastic achievement (house end-of-year grade average)
   12. Efficiency of house organization for the contest (as judged by case readers)
   13. Usefulness of suggestions (as rated by participant-observers)
   14. Contribution to the contest (from end-of-year sociometric schedules)

D. Patterns of leadership:
   15. Strength of leadership (as judged by case readers)
   16. Type of leadership (as judged by case readers)
   17. The leader role of the counselor (as judged by case readers)
   18. Support of formal leader (from end-of-year sociometric schedule)

[3] Definitions of these 18 variables and Table 4 containing their house means and ranks have been filed with the American Documentation Institute. Order Document 3616 from American Documentation Institute, 1719 N Street, N.W., Washington 6, D. C., remitting $1.00 for microfilm (images 1 inch high on standard 35 mm. motion picture film) or $1.20 for photocopies (6×8 inches) readable without optical aid. A longer manuscript of the report also is available for reference from the senior author.

These variables range all the way from clearly objective data (school grades and attendance data for house meetings) to more or less tenuous and possibly unstable measures of interpersonal relations from sociometric schedules. The interactions in each house are, furthermore, seen through the eyes of three different sets of observers: the students themselves who were participating in the contest; the participant-observers, who, with minimal training, were recording house progress during the contest as well as participating; and the two judges who were the only ones to have a view of all participant-observer reports, for purposes of comparing the houses on defined behavior.

Because the basic N for the sample of houses is thirteen, rank-order correlations must be used as the approximate estimate of relationships. Implicit in this procedure is the assumption that the behavior of a particular house can be indexed by averaging behavior measures for its membership of seven to sixteen students taken singly, i.e., that the group is merely the average of its component parts, disregarding the variability of the component members around the group average. An alternative assumption, and one equally crucial to test, is that the group's behavior in certain areas is best indexed by minimum variability around a known average position, as would be the case, for example, in estimating the possibility of change in attitudes for a group showing a strong original attitude in *opposition* to prevailing authority, accompanied by a small standard deviation around the mean opposition score.

A further assumption, more directly amenable to testing in this study, concerns the varieties of behavior defined by the eighteen independent variables. These were defined and grouped on essentially *logical* grounds at various stages of the year's field study. As a first approximation, therefore, the behaviors so defined and grouped should show relatively high patterns of intercorrelations within a group and relatively low correlations with behaviors from other groups, if our formulations are to have meaning. More specifically, any two independent variables which appear to be closely related logically and operationally might be expected to show a sufficiently high intercorrelation to be interpreted as a measure of the reliability of two assessment methods. Where such expectations fail to materialize, either the variables observed derive from different domains of behavior, in spite of the a priori logic by which they were grouped, or the methods of measuring the variables are insufficiently precise, or the variables, while relatively unrelated, may still tap effectively component parts of the criterion measure and might show a substantial loading of a more general factor underlying the total matrix.

## Results

*The Criterion.* Table 1 lists the rankings assigned by each of five faculty judges to the final reports prepared by each of the thirteen houses. Judges 3 and 4 appear to deviate most markedly from the remaining three judges in their evaluations. However, on the basis of the average ranking of the five judges, Houses M, E, D, K, and G, respectively, were awarded the announced cash prizes as soon as possible after the judges had made their independent reports to the research team and before the intercorrelations of judges had been determined.

Table 1

Ranks Assigned by Each Judge to Essays Produced by Specified House Unit*

| | Judge | | | | |
|---|---|---|---|---|---|
| House | 1 | 2 | 3 | 4 | 5 |
| A | 11 | 12 | 6 | 9 | 9.5 |
| B | 2 | 9 | 9 | 5 | 4.5 |
| C | 12 | 11 | 2 | 6 | 11.0 |
| D | 8 | 4 | 1 | 4 | 6.5 |
| E | 4 | 5 | 5 | 2 | 4.5 |
| F | 13 | 13 | 13 | 13 | 13.0 |
| G | 5 | 2 | 4 | 11 | 2.0 |
| H | 10 | 8 | 11 | 10 | 12.0 |
| I | 6 | 6 | 7 | 3 | 8.0 |
| J | 7 | 7 | 12 | 8 | 6.5 |
| K | 3 | 3 | 8 | 7 | 3.0 |
| L | 9 | 10 | 10 | 12 | 9.5 |
| M | 1 | 1 | 3 | 1 | 1.0 |

* The five cash prizes were awarded to Houses M, E, D, K, and G, respectively.

Table 2, showing the rank-order intercorrelations between all pairs of judges, indicates more clearly the degrees of agreement. The average intercorrelation of Table 2, based on five judges, is .59. Averaging the intercorrelations of Judges 1, 2, and 5, the resultant value is .85. Because these three judges showed a greater comparability and consistency in their evaluations, their rankings only have been used in the more detailed analyses of the factors associated with the productivity of the groups.

Another aspect of the criterion should be mentioned briefly. While the residents of the Village participated quite willingly in the task imposed by the experimenters, even to the extent of requesting that their detailed suggestions for improvement be sent anonymously at the end of the contest to the University administration, the task toward which they directed their energies involved essentially a verbal formulation and expression of situations they had self-consciously been experiencing and discussing throughout the year. The better performers in preparing such a report might therefore be the "better" group, on any one of a number of scales of "goodness" as a well-satisfied, well-organized residence unit. Had the members of each house been faced with a task unrelated to their experiences of living together, the resultant criterion measure might show an entirely different set of relations to the same independent variables or the maintenance of the same relative weights at a much lower level of absolute predictive power.

These reservations about the criterion can be viewed in at least two ways. First, one might argue that the results of the study cannot be generalized—that group forces relating to productivity cannot be presumed to operate independently of the content of the task toward which the group is oriented as the criterion measure. Alternatively, one might argue that the predictor variables will retain their relative weights for any dependent (criterion) variable of approximately the same degree of relevance to the group. Relevance, in this sense, becomes the crucial property or quality of the criterion task to be perceived by the participants and transcends the task itself to a degree that will permit generalization. The property of relevance might be compounded of no less than the following elements: the group's perception of the importance of the task; the group's desire to accomplish the task; the group's confidence that its members know how to do the task.

Table 2

Rank-Order Intercorrelations between Pairs of Judges in Evaluating House Essays

| | Judge | | | |
|---|---|---|---|---|
| Judge | 2 | 3 | 4 | 5 |
| 1 | .78 | .23 | .61 | .92 |
| 2 | | .47 | .52 | .85 |
| 3 | | | .59 | .42 |
| 4 | | | | .53 |

Average rho, 5 judges = .59. Average rho, Judges 1, 2, and 5 = .85.

*The Predictor Variables.* All house averages on the criterion variable, and also on the eighteen predictor variables previously described, were obtained and used in arriving at the rank-order correlations set forth in Tables 3 and 4.[4] By reference to the list of variables given earlier, it is possible to study the categories of predictor variables and to make some observations regarding the consistency with which certain behaviors were observed in this study.

Variables 4, 10, 12, 15, 16, and 17 in Table 3 resulted from the average ratings of two independent judges, who read all the participant-observer reports and rated them for the variables involved. The degree of agreement between these two judges in assigning house ratings to these reports is shown in Table 5.

In assessing strength of leadership, efficiency of organization for the task, and degree of enthusiasm of the participants, the judges are in substantial agreement, although they show less agreement for the other three variables.

Variables 2, 3, and 13 are derived from ratings made after each meeting by the participant-observers, who rated each girl in attendance, including themselves, on a five-step scale

[4] Table 4, giving house averages, and resultant ranks, has been filed with the American Documentation Institute. See footnote 3.

Table 3

Rank-Order Intercorrelations among 18 Independent Variables and the Dependent Variable of Rated Productivity

| Variables | Productivity | Variables | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 1. Volume of participation | .02 | | | | | | | | | | | | | | | | | |
| 2. Extent of participation | .21 | .26 | | | | | | | | | | | | | | | | |
| 3. Assumption of responsibility | .25 | −.21 | .74 | | | | | | | | | | | | | | | |
| 4. Enthusiasm for the task | .59 | .43 | .66 | .51 | | | | | | | | | | | | | | |
| 5. Satisfaction with Village life | .64 | .30 | −.03 | −.09 | .27 | | | | | | | | | | | | | |
| 6. Dislike ratio | .38 | −.33 | .17 | .16 | .38 | .16 | | | | | | | | | | | | |
| 7. In-group/out-group: "social support" | .50 | −.35 | .29 | .18 | .19 | −.14 | .37 | | | | | | | | | | | |
| 8. In-group/out-group: "confidantes" | .19 | .10 | .16 | .19 | .11 | .49 | .04 | −.03 | | | | | | | | | | |
| 9. In-group/out-group: "intimate friends" | .27 | .08 | −.45 | −.45 | −.27 | .57 | −.37 | 0 | .34 | | | | | | | | | |
| 10. Evidence of personality conflict | .25 | −.04 | .17 | .17 | .52 | −.22 | .58 | .32 | .34 | −.46 | | | | | | | | |
| 11. House scholastic achievement | −.01 | −.02 | .19 | .52 | .27 | −.25 | −.23 | −.02 | .39 | −.16 | .09 | | | | | | | |
| 12. Efficiency of house organization for the contest | .86 | .03 | .48 | .45 | .83 | .46 | .56 | .41 | .10 | −.04 | .46 | .07 | | | | | | |
| 13. Usefulness of suggestions | .13 | .27 | .80 | .77 | .72 | −.12 | .11 | .12 | .27 | −.50 | .37 | .65 | .42 | | | | | |
| 14. Contribution to the contest | −.05 | .41 | .46 | .23 | .35 | .09 | .003 | −.15 | .24 | −.21 | .12 | .12 | .11 | .54 | | | | |
| 15. Strength of leadership | .16 | −.33 | −.13 | −.07 | −.13 | −.11 | .007 | .45 | .24 | .34 | −.27 | .10 | .02 | −.25 | −.26 | | | |
| 16. Type of leadership | .40 | .15 | .24 | .03 | .40 | .34 | .15 | .10 | −.08 | .30 | .18 | −.31 | .54 | .02 | .37 | .16 | | |
| 17. Leader role of the counselor | .47 | −.39 | −.30 | −.07 | .04 | −.003 | .14 | .61 | .02 | .38 | .23 | .22 | .28 | −.13 | −.11 | .60 | .19 | |
| 18. Support of formal leader | .51 | .30 | .55 | .37 | .83 | .35 | .63 | .17 | .39 | −.24 | .45 | .15 | .73 | .54 | .24 | −.06 | .31 | −.06 |

Table 5

Rank-Order Correlations of Ratings of Participant-Observers' Reports by Two Judges (N = 13)

| Variable | Rho |
|---|---|
| Enthusiasm for task | .84 |
| Evidence of personality conflict | .48 |
| Efficiency of house organization | .81 |
| Strength of leadership | .89 |
| Type of leadership | .51 |
| Leader role of counselor (also participant-observer) | .46 |

for the behavior in question. It is evident from the high intercorrelations of these three variables in Table 3 that the participant-observers seem to be working within the framework of a general halo about the girls under their observation, with insufficient specificity.

Variable 14 represents the judgment of the actual participants in each house regarding their own contributions and the contributions of their house-mates in the task of preparing the report. This variable might be expected to show a marked degree of agreement with the judgments of the participant-observers, as indexed in Variables 2, 3, and 13. In looking at the three relevant correlations, however, it would not appear that the participants themselves and the participant-observers were in close agreement on the behavior under consideration. Nor does the girls' evaluation of each other's contributions relate closely to the ratings of the judges who saw the progress of the contest through the narrative reports submitted by the participant-observers (Variable 14 vs. Variables 4, 10, 12, 15, 16, and 17).

There does not appear to be extremely high agreement among the participants, the participant-observers, and the judges who reviewed the entire range of participant-observer reports in the assessment of behaviors that on a priori grounds appear to be related. Whether this may be due to poor training of the raters, disjunctions between relative and absolute judgments, or psychological differences between apparently similar behaviors cannot be determined from the data. It is likely that all three of these factors contribute to the low patterns of intercorrelations.

Table 3 also permits an analysis of the communality of the groups of predictor variables. Under the general rubric of "acceptance of the group goal," for example, four specific variables are included: volume of participation; extent of participation; assumption of responsibility; and enthusiasm for the task. The six possible intercorrelations among these four variables appear near the top of Table 3. The average intercorrelation for this set of six, including the negative correlation, is approximately .40—a correlation that does not warrant any assumption of great communality among the variables logically grouped as facets or coordinates of the general concept of "acceptance of the group goal." Furthermore, these four variables show a wide range of correlation with the criterion measure; only one of the four—enthusiasm—approximates a relationship significantly greater than zero for the N of 13.

Similar statements can be made about the other groups of predictor variables, with respect to their patterns of intercorrelations as well as their correlations with the criterion variable.

These data have some bearing on the conceptualization phase of designing experiments on small groups, since the concept is in part given meaning by the choice of behavior deemed to be coordinate with it. In Table 3, for example, six possible behavioral indices of "satisfaction with previous group experience" are presented, and the intercorrelations among them are shown. In addition, when each is correlated with the criterion, only two (satisfaction with Village life and ratio of in-group to out-group choices for "social support" on the first sociometric schedule) approximate significance in predicting productivity with the N of 13. A posteriori, some of the attenuating factors become evident: the six behavioral indices are drawn from different time points in the experimental year; they are assessed or measured with indeterminate accuracy and by different observers; they are drawn from behavioral domains, in terms of low intercorrelations of sociometric data, showing considerable and surprising specificity (**2**).

The problem of identification, accurate assessment, and meaning of variables, is certainly

not a new one in psychology; it applies with equal cogency in research on groups as well as research on individuals and it emphasizes the importance of careful derivations of behavioral coordinates in testing the relationships involved in group processes.

*The Predictions of Productivity.* Table 3 permits the choice of several possible combinations of variables for use in a multiple regression equation, in such a way that the earlier hypothetical formulation can be supported with varying degrees of confidence. For example, the Wherry-Doolittle shrinkage formula[5] can be used to estimate the *R* with productivity that might be obtained from a multiple containing: volume of participation; satisfaction with Village life; achievement; and percentage of votes received by the formal leader. These four variables commend themselves as relatively objective measures of factors crucial to the test of the general formulation of the problem as given at the start of this paper. Two of these four give as an estimate of *R* a value of .676; these two are: satisfaction with Village life and percentage of votes received by the formal leader. The addition of the other two variables shrinks the results to a lower estimate of *R*.

But the following three variables give an estimated *R* of .867: satisfaction with Village life; percentage of votes received by the formal leader; and ratio of in-group to out-group choices for "social support" on the sociometric questionnaire given at the start of the school year. Note that two of these three variables are drawn from the same category of variables, dealing with "satisfaction with previous group experiences."

An estimated *R* of .88 is obtained by use of the following four variables: enthusiasm as estimated by the judges who read the participant-observers' reports; evidence of personality conflict as seen by the same judges; efficiency of house organization for the contest; and amount of leadership judged to have been exercised by the counselor.

Another estimated *R* has been established on still another combination of variables, each of which is drawn from one of the four general rubrics of the original hypothetical formulation and each of which shows a substantial correlation with the criterion of productivity: enthusiasm for the task; satisfaction with Village life; efficiency of organization for the task; and percentage of votes received by the formal leader. The *R* value is .896. When an estimated *R* is computed using only three of these four variables (satisfaction, efficiency and formal leadership support) the multiple *R* becomes .904.

It appears that in this particular experimental setting and with the criterion task here defined, it is possible to identify factors contributing to the relative productivity of the thirteen groups. The estimated *R*'s will differ according to the variables chosen for inclusion in the formula; in one instance, at least, a set of variables not on a priori grounds completely consonant with the conceptualization yields a higher estimated *R* than sets of variables more closely related to the hypotheses on logical grounds. However, since one variable alone—efficiency of house organization—carries the greatest predictive weight, the multiple regression solution is somewhat uneconomical here. The study has raised more problems than it has solved, one of which warrants mention. The multiple regression solution assumes a linear and additive relation between the independent variables and the criterion; it requires also that enough independent variables be included to account for all or almost all criterion components. Even if, in replications, the criterion components have varying weights for other samples of groups, the beta weights of the independent variables should still remain relatively constant. It is essential, however, to consider whether productivity is a linear and additive resultant of various group factors that can be hypothesized, or whether alternative formulations are closer to the reality of group behavior.

If linearity is assumed, a possible next step in the present study would be the factorial reduction of the variables represented in Table

[5] Stead, W. H., Shartle, C. L., *et al. Occupational counseling techniques.* New York: American Book Company, 1940, pp. 245–51. The assumption that rank-order correlations are direct equivalents of Pearsonian values has been made in using this formula, even though the authors realize it is open to question, since the problem here is the *relative* changes in *R* estimates from alternative selections of variables in the same matrix.

3, including the criterion variable. This solution, however, would be meaningful only to the extent that one is dealing with reliably estimated and conceptually relevant behaviors of a sample of groups known to be representative of some population of groups, and to the further extent that the criterion is truly a meaningful task for the groups involved.

As a result of discussions of such problems of group productivity, Schachter set up a more incisive experiment on the relations between cohesiveness and effect of group induction to maximize or restrict group productivity (**3**). His study should be read in connection with the present report, since it introduces a variable not dealt with here—differential effect of group induction in high and low cohesive groups. In brief, he finds no necessary relationship between cohesiveness and high productivity. Group members will accept induction either to increase or decrease production. Cohesiveness appears to be a determining variable in the "slow-down" condition but not in the "speed-up" condition. When the group induces forces to increase production, both high and low cohesive subjects accept group induction and increase their output markedly. When the group induces forces to decrease production, the high cohesive subjects are more accepting of group induction and consequently less productive than low cohesive subjects. While the designs of these two studies are different, they deal with the same general problem and together may form the basis for further research on group productivity.

*Received January 5, 1952.*

## References

1. Darley, J. G., Gross, N., and Martin, W. E. Studies of group behavior: stability, change, and interrelations of psychometric and sociometric variables. *J. abnorm. soc. Psychol.*, 1951, **46**, 565–576.
2. Martin, W. E., Darley, J. G., and Gross, N. Studies of group behavior: methodological problems in the study of interrelationship of group members. *Educ. psychol. Measmt.*, 1952, **12**, Winter Number.
3. Schachter, S., Ellertson, N., McBride, Dorothy, and Gregory, Doris. An experimental study of cohesiveness and productivity. *Hum. Relat.*, 1951, **4**, 229–238.

# MMPI Personality Patterns for Various College Major Groups

**Ralph D. Norman and Miriam Redlo**

*The University of New Mexico*

A recent study by Daniels and Hunter (**3**) raises two questions which are important in guidance, namely, whether there are personality patterns which gravitate toward certain occupations, and also whether there are fixed "personality demands" in various jobs. It is the purpose of the present research to extend these questions to the college student and his choice of major subject. Do students who select a major in business, say, have certain personality characteristics in common and are they different from those who choose art or psychology? It was decided to use the MMPI to help answer these questions since it has validity in pointing out vocational personality trends (**1, 8, 12, 22, 23**), although Lough (**13, 14**) expresses doubt about this.

## Subjects and Procedure

Subjects were 149 male seniors and graduate students at The University of New Mexico, chosen on the assumption that if personality is related to choice of major, the advanced students would exhibit the most pronounced characteristics. They were distributed as follows: psychology, 17; sociology, 3; mathematics, 2; chemistry, 4; physics, 12; engineering, 29; anthropology, 22; business administration, 23; art, 16; music, 1; geology, 8; biology, 4; physical education, 2; history, 3; English, 1; elementary education, 1; Inter-American Affairs, 1. These majors were grouped together because of logical relationships into seven major groupings: psychology-sociology (PS), 20; mathematics-chemistry-physics (MCP), 18; engineering (Engr), 29; anthropology (Anth), 22; business administration (BA), 23; art-music (AM), 17; geology (Geol), 8. The remainder constituted a miscellaneous grouping which was not contrasted with the others because of its heterogeneity, except that it was included in the total group when each grouping was contrasted with the total minus itself.

All *Ss* took the MMPI, and completed a questionnaire, rating their degree of satisfaction with their major on a 7-point scale and indicating which major they would choose if they could reselect it. These latter data were used because it was hypothesized that those who indicated greatest satisfaction would exhibit the most definite or pronounced personality pattern for their own grouping. Similarly, those rechoosing the same or similar majors, it was felt, should be more like their own grouping than those rechoosing differently.

Results from the attitude scale revealed that it differentiated in numbers largely between "strongly satisfied" and "satisfied" or less, probably because of natural reluctance to admit dissatisfaction with major after one had spent considerable time and effort with it. Data for all 149 cases showed the following distribution: strongly dissatisfied, 0; dissatisfied, 0; mildly dissatisfied, 3; neutral, 4; mildly satisfied, 14; satisfied, 66; strongly satisfied, 62. Consequently, it was decided to compare the strongly satisfied (hereafter SS) groups within each major against the satisfied-and-less (hereafter LS).

Also, data on rechoice of major were analyzed into 3 categories: "different" (e.g., a student might indicate a rechoice of a completely different major, as from Engr to BA); "different but similar" (e.g., a rechoice of chemistry instead of physics); and "same." Two "undecided" cases were not used, nor was the Misc grouping, the reason therefor being evident below.

By use of the *t* test, the mean score of each major grouping was compared with the mean score of every other major grouping on all MMPI Scales (except ? Scale); also, the mean score of each grouping was compared with the mean of the total group less the grouping in question. Similarly, the means of SS groups were contrasted with those of LS for each major grouping and for the total group. Statistical analysis of the data concerned with rechoosing majors was more difficult since the numbers in the categories (other than "same") were very small. Hence it was decided to use the mean T-score deviation on each scale from the mean T-score of the related major grouping. The trend of deviations might then suggest whether or not those who would rechoose different majors deviate more markedly from the means of their major groupings than those who would reselect the same major.

## Results

1. *Results for Major Groupings Compared with Each Other and with "Total."*[1] Table 1 presents the means and sigmas of T-scores of

[1] "Total" refers to the total group of 149 minus the particular subgrouping.

Table 1

Means and Sigmas of *T*-Scores of Each Major Grouping and Total Group on the MMPI

| Major Grouping | *N* | | *L** | *F** | *K* | *Hs* | *D* | *Hy* | *Pd* | *Mf* | *Pa* | *Pt* | *Sc* | *Ma* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Psychology and Sociology | 20 | M | 2.4 | 4.5 | 60.3 | 53.7 | 51.8 | 58.9 | 61.1 | 61.2 | 48.7 | 55.1 | 59.0 | 61.5 |
| | | σ | 1.2 | 3.1 | 8.7 | 5.3 | 9.8 | 6.7 | 5.4 | 9.4 | 7.1 | 7.6 | 7.7 | 11.3 |
| Mathematics, Chem. and Physics | 18 | M | 3.4 | 4.2 | 55.8 | 54.9 | 58.2 | 56.9 | 55.0 | 61.9 | 53.3 | 55.8 | 52.8 | 55.8 |
| | | σ | 2.2 | 2.3 | 6.4 | 7.1 | 10.8 | 6.7 | 9.6 | 9.4 | 7.2 | 8.5 | 8.0 | 9.8 |
| Engineering | 29 | M | 2.3 | 3.3 | 56.3 | 53.1 | 54.3 | 55.3 | 55.2 | 59.2 | 54.9 | 55.8 | 54.7 | 54.6 |
| | | σ | 1.7 | 2.7 | 8.4 | 8.7 | 10.9 | 7.1 | 11.5 | 10.2 | 6.7 | 11.7 | 10.4 | 10.3 |
| Anthropology | 22 | M | 3.0 | 5.5 | 57.1 | 53.0 | 55.1 | 56.8 | 55.0 | 64.7 | 54.6 | 54.6 | 55.5 | 58.3 |
| | | σ | 1.7 | 3.0 | 6.3 | 8.6 | 8.9 | 6.5 | 10.0 | 7.1 | 7.4 | 8.2 | 10.4 | 11.9 |
| Business Administration | 23 | M | 3.1 | 4.4 | 58.0 | 56.1 | 49.4 | 58.5 | 59.0 | 59.8 | 54.2 | 57.8 | 57.0 | 60.8 |
| | | σ | 2.2 | 2.9 | 10.4 | 9.6 | 6.8 | 7.3 | 12.3 | 9.1 | 9.1 | 8.7 | 9.8 | 11.2 |
| Art and Music | 17 | M | 3.8 | 4.6 | 61.0 | 55.9 | 55.1 | 61.1 | 60.8 | 72.7 | 56.9 | 57.6 | 61.7 | 62.1 |
| | | σ | 2.0 | 2.8 | 8.1 | 11.4 | 11.3 | 10.5 | 9.7 | 8.6 | 4.6 | 8.1 | 8.6 | 8.4 |
| Geology | 8 | M | 2.6 | 5.1 | 61.3 | 52.4 | 51.8 | 61.0 | 62.6 | 64.0 | 54.1 | 59.4 | 60.8 | 63.4 |
| | | σ | 1.7 | 3.1 | 3.3 | 7.7 | 10.2 | 7.9 | 8.6 | 6.9 | 10.7 | 11.7 | 11.0 | 8.9 |
| Total Group (includ. 12 Misc. Majors) | 149 | M | 2.8 | 4.6 | 57.6 | 54.1 | 53.9 | 57.8 | 58.1 | 62.7 | 53.4 | 56.8 | 57.2 | 59.0 |
| | | σ | 1.9 | 3.1 | 9.4 | 8.7 | 10.5 | 7.8 | 10.4 | 9.9 | 7.8 | 10.2 | 10.5 | 10.0 |

* Raw scores.

each major grouping and the total group of 149 on the MMPI.[2] Table 2 lists the significant differences, checked for homogeneity of variance, among the major groupings, and between the particular major grouping and the "total" group. Significant differences, discussed later, appear for all scales except *Pt* and *Hs*.

2. *Results for SS Compared with LS.* No significant differences appeared when means for these 2 groups were compared *within* each grouping. The small *N*'s and perhaps our manner of dichotomizing the groups may have affected this approach. Accordingly another approach was tried. In Table 2, 8 instances occurred wherein the scales discriminated between the particular major grouping and the "total." It might then be hypothesized that the SS group within each major should have a mean score in relation to (i.e., higher or lower than) that of the LS group similar to that of the major grouping itself in relation to "total." Table 3 gives results of this contrast. According to Table 3, the prediction holds in 5 cases, falls down in 2 and is indecisive in 1 (*Pa* for PS) where the two means are almost identical. The best conclusion we may make here is that there is a tendency for SS groups to be more like their own major groupings on certain discriminative scales than LS groups.

When the total group of 149 students was divided into SS and LS, 2 statistically significant MMPI differences emerged. Regardless of major, the 65 in the SS group exhibit *more* feminine interests (*Mf* mean 65.02) than do the 87 in the LS group (*Mf* mean 61.01), the difference being significant at the 2 per cent level ($t$ of 2.42). In addition, the SS group was significantly less psychasthenic (*Pt* mean 54.85) than the LS group (*Pt* mean 58.23), the difference being significant at the 5 per cent level ($t$ of 2.17). The SS group also had a significantly higher major grade point average (2.23) than the LS group (1.79), the $t$ being 4.40, $P < .001$.

3. *Results Concerned with Rechoosing Majors.* As noted above, it was possible, when students were asked about rechoice, to separate each grouping into 3 categories. In view of smallness of *N*'s, no statistical technique was em-

[2] Except for *L* and *F* Scales which are reported in raw score form since Hathaway and Meehl (11) state the original *T*-score tables for these were inaccurate.

Table 2

Significant Differences (*t*'s) between Major Groupings on the MMPI

| | *L* | | *F* | | *K* | | *D* | | *Hy* |
|---|---|---|---|---|---|---|---|---|---|
| AM> | PS, 2.62*<br>Engr, 2.78**<br>Tot, 2.27* | Engr< | Anth, 2.58*<br>Tot, 2.54* | MCP< | AM, 2.05*<br>Geol, 2.19* | BA< | MCP, 3.11**<br>Anth, 2.37*<br>Tot, 2.99**† | AM>Engr, 2.19* | |
| | *Pd* | | *Mf* | | *Pa* | | *Sc* | | *Ma* |
| PS> | MCP, 2.30*†<br>Engr, 2.35*†<br>Anth, 2.41*† | AM> | PS, 3.75**<br>MCP, 3.43**<br>Anth, 3.07**<br>BA, 4.41**<br>Geol, 2.39*<br>Engr, 4.38**<br>Tot, 4.71** | PS< | Engr, 3.11*<br>Anth, 2.58*<br>BA, 2.15*<br>AM, 4.84**<br>Tot, 3.34** | MCP< | PS, 2.30*<br>AM, 3.07**<br>Tot, 1.96* | Engr< | PS, 2.19*<br>BA, 2.04*<br>AM, 2.45*<br>Geol, 2.15*<br>Tot, 2.71** |
| | | Anth>Engr, 2.12*<br>Engr<Tot, 2.09* | | | | AM>Engr, 2.26* | | | |

Tot = Total group minus contrasted subgroup.
* Significant at .05 level.
** Significant at .01 level.
† Corrected for heterogeneity of variance.

ployed to test for significance but in 5 of 7 instances (all except Anth and Engr), those students who would select different majors if given the opportunity to choose again deviated more from their own major grouping than those who would rechoose the same major. The trend is perhaps more pronounced if the total group of 135 who made a definite rechoice is considered. Mean T-score deviation from major grouping is greatest for "different" ($N = 13$), $M = 7.55$; next greatest for "different but similar" ($N = 8$), $M = 7.28$; and least for "same" ($N = 114$), $M = 6.60$. These results must be considered suggestive rather than conclusive of the fact that personality needs exist in academic choice.

Table 3

Strongly Satisfied (SS) Compared with Satisfied-and-Less (LS)

| Prediction from Table 2 | | SS | | LS | | Prediction Verified? |
|---|---|---|---|---|---|---|
| | | *N* | *M* | *N* | *M* | |
| *L* | AM>"total" | 14 | 3.4 | 3 | 5.7 | No |
| *F* | Engr<"total" | 5 | 2.2 | 24 | 3.6 | Yes |
| *D* | BA<"total" | 8 | 48.3 | 15 | 50.1 | Yes |
| *Mf* | AM>"total" | 14 | 73.6 | 3 | 68.3 | Yes |
| *Mf* | Engr<"total" | 5 | 56.2 | 24 | 59.9 | Yes |
| *Pa* | PS<"total" | 7 | 48.7 | 13 | 48.6 | ? |
| *Sc* | MCP<"total" | 4 | 57.0 | 14 | 51.6 | No |
| *Ma* | Engr<"total" | 5 | 53.8 | 24 | 54.7 | Yes |

## Discussion

Discussion of results will not be concerned with explaining every significant difference in Table 2, for when 336 different *t*'s are run, it is possible that some may reach a level of significance by chance. Rather, to be conservative, it would be more pertinent to discuss *trends* evidenced in Table 2 where significant differences were found at least 3 times between a grouping and other groupings or between a major grouping and the total.

1. *Lie Scale.* On this scale, AM is significantly *greater* than PS, Engr, and "total." Although this scale has not been analyzed clinically, Hathaway and McKinley (**10**) state that high *L* results when the subject chooses responses which make him most socially acceptable. It is of interest here that Grayson (**7**) points out that high *L* is usually accompanied by high *Hy*. The AM grouping shows

both highest *L* and highest *Hy*, and on both scales differs significantly from Engr which has lowest *L* and *Hy*. Lying and hysterical reactions are both socially oriented, implying sensitivity for applause of the crowd, a need which is probably greater in artists than engineers (e.g., artists put their products on public display). Moreover, the artist converts reality into fantasy, and from his earliest training precise regard for realism is not encouraged. This is quite the opposite of Engr's training.

2. *F Scale.* On this scale, Engr is significantly *lower* than "total." Hathaway and McKinley (**10**) contend that high *F* may be due to carelessness and low *F* indicates that *S*'s responses were rational and relatively pertinent. Engineering is a precise, careful, rational field, completely pertinent to practical matters. It is indeed possible that the low *F* of Engr is a carry-over of his training. It is of interest here that Cook and Wherry (**2**) found a Mechanical Coordination Factor, resulting from a factor analysis of MMPI and aptitude tests, had a significant positive loading on *F*. They state that subjects with good mechanical coordination were meticulous in answering personality questions.

3. *D Scale.* BA is significantly *lower* than MCP, Anth, AM, and "total." Our results are similar to those of Paterson (**15**), Trabue (**21**), and Dodge (**5**, **6**) who found sales-people highly socially dominant. Harrower and Cox (**9**) found insurance salesmen giving evidence of considerable drive. As compared to other groups, the presence of depressive trends in BA would be a definitely serious liability.

4. *Pd Scale.* PS scores significantly *greater* than MCP, Engr, Anth, and "total." It seems paradoxical that PS should have a higher *Pd* implying a tendency toward psychopathic characteristics. Perhaps the best explanation is rather obvious. In the psychopath, freedom of expression is paramount—he defies convention and tradition. Since sociology and psychology attempt to blame many of the present taboos for mental illness and social problems, it is possible that the PS group feels that it must defy convention. Or perhaps as a result of training PS does not feel any reluctance about answering *Pd* items truthfully or may even interpret them differently. These attitudes would be picked up by an inventory like the MMPI.

5. *Mf Scale.* AM is significantly *greater* than all other groupings and "total." A valid explanation would be that sensitive or cultural work has been characterized as feminine, and since art requires both, AM will naturally exhibit higher *Mf*. Roe (**17**), studying artists, states that they have a type adaptation which is decidedly non-aggressive, hence more feminine than masculine. However, she does state further that passivity is more characteristic of intellectual males; this statement is in line with our finding that *Mf* is the highest of all for our total group of 149. On this scale also Engr scores significantly lower than "total," a not unexpected result in view of the "masculinity" of engineering. Cook and Wherry (**2**) found a negative loading on mechanical aptitude for those with high *Mf*.

6. *Pa Scale.* PS scores significantly *lower* than all other groups (except Geol), and "total." It seems that no explanation in terms of personality is required. The sophistication of PS enables it to select out the *Pa* questions, especially in view of the fact that the paranoid is a dramatic type.

7. *Sc Scale.* MCP is significantly lower than PS, AM, Geol, and "total." It is probably true that the reality of MCP represents a rather unique reality as compared to less pure and emotionally rarefied fields. The nature of this reality is in no way comparable, however, to schizoid reality. The former is logical, rational, intellectual and highly abstract, while the latter is bizarre, irrational, emotional and concrete.

8. *Ma Scale.* Engr scores significantly lower than PS, BA, AM, Geol, and "total." Daniels and Hunter (**3**) state that the work needs of the *Ma* pattern are related to an outlet for enthusiasm and a high degree of overt activity. We feel that both MCP and Engr require the least overt activity in comparison to other groups; they do not require as much social orientation as PS, e.g., or as strong an outlet for enthusiasm as AM, BA, or Geol. The engineer, and also MCP, have a definite concept of reality and their place in it, and therefore, unlike the hypomanic individual they do not shift their attention constantly. However, our results on Engr and the *Ma*

Scale appear at variance with a finding by Cook and Wherry (**2**). They found a factor labeled Tendency to Over-Activity with positive loadings on mechanical aptitude and knowledge, and state that over-active individuals find outlet in mechanical pursuits. However, the difference between their subjects, naval enlisted submarine candidates, and ours may explain this discrepancy.

An attempt may now be made to describe tentatively each academic group in terms of personality characteristics or of personality demands of the major.

1. *PS.* This group creates somewhat of a problem since their knowledge of psychological dynamics may have contaminated their profiles. The MMPI indicates that, as a group, PS may be characterized by fairly strong *Pd* tendencies and corresponding *Ma* behavior. Roughly, their orientation is outgoing with expressed disregard for convention.

2. *MCP.* This group indicates a pattern of tendencies away from *Pd* and *Ma*, with somewhat high *D* and very low *Sc*. Justification for this pattern may be found in Roe's Rorschach work on scientists. She believes that men with dominantly non-verbal ability, who seem to find satisfaction in empathy with symbols of things rather than things themselves, are men who are suited for and attracted to science (**19**). Therefore, for MCP, the conclusion might be suggested that very weak *Pd*, *Ma*, and *Sc* tendencies are necessary for success since these sciences are concerned with symbols, ideas, and abstraction, rather than with people. The exacting nature of these sciences may contribute to the presence of somewhat higher *D*, a finding similar to that of Roe (**20**), who reports physicists with a high degree of anxiety because of the nature of their work.

3. *Engr.* Tendencies similar to MCP are found in Engr. This group displays very low *Pd*, *Ma*, and *Hy*. Their *Sc* is a little higher than MCP, but lower than the other groups. Explanations similar to those given for MCP appear to apply to engineers, except that they are the most "masculine" (lowest *Mf*), a finding to be expected in view of their mechanical interests. The personality demands of this group seem to imply a stability of behavior, low overt activity, reduced interest in people, and masculinity.

4. *Anth.* This group does not indicate strong patterns. Anth has a relatively high *Mf* score which could mean that this field requires a high degree of sensitivity, a finding in agreement with that of Roe (**18**) in her study of vertebrate paleontologists.

5. *BA.* As a group, BA may be characterized by low *D*, accompanied by *Ma* tendency and social aggression, a factor resulting from the comparatively low *Mf* score. It is obvious that the business world demands the opposite of what McKinley and Hathaway (**10**) say about high *D*, namely, "poor morale . . . feeling of uselessness . . . inability to assume normal optimism." Business also demands aggressive tendencies which should show up in lower *Mf*.

6. *AM.* This group is represented by a complexity of characteristics. Extremely strong feminine sensitivity (high *Mf*) is accompanied by high *Hy*, *Sc*, *Pd*, *Ma*, and somewhat *Pa* tendencies. The generally high mean scores of AM suggest that, as a group, it is somewhat more maladjusted than the others, and may find outlets through artistic sublimations or compensations as suggested by Prados (**16**). In view of the complex pattern of AM, it is difficult to state any personality demands for it, except the high femininity.

Because of the small *N* in Geol, no attempt will be made to define personality characteristics or demands. However, these few cases are not similar to MCP, but we cannot conclude whether or not this is an artifact of our sampling.

This study also noted the fact that SS students were significantly higher in *Mf* than LS. Since it has been demonstrated before (**4**) that college males have *Mf* scores higher than average males, this result is not surprising. Roe (**17**) states that feministic adaptation is characteristic of the sensitive intelligent male, and that intellectual pursuits become a refuge for men who deviate from the pattern of masculinity. College work would thus be more satisfying to the somewhat more feminine male.

We have also noted that LS scores significantly higher than SS on *Pt*. Hathaway

and McKinley (**10**) state that a *Pt* tendency may be manifested via mild depression, excessive worry, lack of confidence, and inability to concentrate. These characteristics should certainly be less prominent among SS students than among LS.

Since ours is an exploratory study, with small *N*'s, we cannot suggest that patterns found here be used for guidance purposes. Research on a much larger scale is needed in order to confirm and expand these findings.

## Summary

Seven groupings of students majoring in different academic subjects were contrasted with each other and with a "total" grouping minus the particular subgrouping on the MMPI. They also rated their satisfaction with their major subject and mentioned their choice of major if allowed to rechoose. Principal findings were:

1. The MMPI is valid for distinguishing personality trends amongst various major groupings. Certain scales significantly discriminated major groupings from the remainder of the students.

2. There is a tendency for students who are strongly satisfied with their major to resemble their own groupings on discriminative scales.

3. Significant differences were found between strongly satisfied and satisfied-and-less students on *Mf* and *Pt*. The former were higher on *Mf*, the latter on *Pt*.

4. When mean T-score deviations from average T-scores are calculated, there is a tendency for students who would rechoose the same major to deviate less from their own groupings than those who would rechoose a different major.

*Received December 24, 1951.*

## References

1. Chyatte, C. Personality traits of professional actors. *Occupations*, 1949, **27**, 245–250.
2. Cook, E. B., and Wherry, R. J. A factor analysis of MMPI and aptitude test data. *J. appl. Psychol.*, 1950, **34**, 260–266.
3. Daniels, E. E., and Hunter, W. A. MMPI personality patterns for various occupations. *J. appl. Psychol.*, 1949, **33**, 559–565.
4. DeCillis, Olga E., and Orbison, W. D. A comparison of the Terman-Miles M-F Test and the *Mf* Scale of the MMPI. *J. appl. Psychol.*, 1950, **34**, 338–342.
5. Dodge, A. F. Social dominance and sales personality. *J. appl. Psychol.*, 1938, **22**, 132–139.
6. Dodge, A. F. What are the personality traits of the successful salesperson? *J. appl. Psychol.*, 1938, **22**, 229–238.
7. Grayson, H. M. *The Minnesota Multiphasic Scales*. Mimeographed, no date.
8. Harmon, L. R., and Wiener, D. N. Use of the MMPI in vocational advisement. *J. appl. Psychol.*, 1945, **29**, 132–141.
9. Harrower, G. J., and Cox, K. J. The results obtained from a number of occupational groupings on the professional level with the Rorschach group method. *Bull. Canad. Psychol. Assn.*, 1942, **2**, 31–33.
10. Hathaway, S. R., and McKinley, J. C. *Manual for the Minnesota Multiphasic Personality Inventory*. N. Y.: Psychol. Corp., 1943.
11. Hathaway, S. R., and Meehl, P. E. *An atlas for for the clinical use of the MMPI*. Minneapolis: U. Minn. Press, 1951.
12. Lewis, J. A. Kuder Preference Record and MMPI scores for two occupational groups. *J. consult. Psychol.*, 1947, **11**, 194–201.
13. Lough, Orpha M. Teachers college students and the MMPI. *J. appl. Psychol.*, 1946, **30**, 241–247.
14. Lough, Orpha M. Women students in liberal arts, nursing, and teacher training curricula and the MMPI. *J. appl. Psychol.*, 1947, **31**, 437–445.
15. Paterson, D. G. Research studies in individual diagnoses. *Univ. Minn. Bull. Emplyt. Stab. Res. Inst.*, 1934, **3**, No. 4.
16. Prados, M. Rorschach studies on artists. *Rorschach Res. Exch.*, 1944, **8**, 178–183.
17. Roe, Anne. The personality of artists. *Educ. psychol. Measmt.*, 1946, **6**, 401–408.
18. Roe, Anne. A Rorschach study of a group of scientists and technicians. *J. consult. Psychol.*, 1946, **10**, 317–327.
19. Roe, Anne. Personality and vocation. *Trans. N. Y. Acad. Sci.*, 1947, Series II, **9**, 257–267.
20. Roe, Anne. Analysis of group Rorschachs of physical scientists. *J. proj. Tech.*, 1950, **14**, 385–398.
21. Trabue, M. R. Occupation ability patterns. *Personnel J.*, 1933, **11**, 344–351.
22. Verniaud, Willie M. Occupational differences in the MMPI. *J. appl. Psychol.*, 1946, **30**, 604–613.
23. Wiener, D. N., and Simon, S. Personality characteristics of embalmer trainees. *J. appl. Psychol.*, 1950, **34**, 391–393.

# The Effects of Tachistoscopic Training in an Adult Reading Program [1]

**George Manolakes**

*Division of Research, State Education Department, University of the State of New York*

The role of the tachistoscope in a reading improvement program has been the subject of much discussion. Since the tachistoscope has been one of the training devices used in the Reading Improvement course at the Marine Corps Supply Schools, this study was initiated in order to determine the effects of omitting tachistoscopic training from the course of instruction.

## Design of Experiment

Accordingly, a study was designed which would provide for an Experimental and a Control group to undergo training in reading improvement simultaneously. The variable element within their instruction was the exclusion of tachistoscopic training from the program of the Experimental group, and the extension of instruction in vocabulary and comprehension skills.

The effects of the variable courses of instruction were measured in terms of the changes that would normally be expected from training with tachistoscopic devices. The Ophthalmograph was selected as the means by which certain of the visual skills might be measured quantitatively. Each student was tested with the Ophthalmograph prior to instruction and at the conclusion of the training period. The film records of the eye movements were utilized in determining: (a) the mean number of fixations; (b) the mean number of regressions, (c) the mean span of recognition; and (d) the mean duration of fixations.

In addition to the changes in visual skills, changes in reading rate have also been attributed to tachistoscopic training. As an index of reading rate, the *Silent Reading Check RRRP G8-5* was administered prior to instruction and the *Silent Reading Check RRRP G8-6* was used at the conclusion of the program. The data were limited to reading rate since it was felt that this would be a more appropriate index since tachistoscopes are more related to the perceptual skills rather than those in comprehension.

*Population.* The population to be studied was a group of Marine Corps officers under instruction at the Marine Corps Supply Schools. The group of 34 officers included 1 captain, 8 first lieutenants, 24 second lieutenants, and 1 commissioned warrant officer.

The officers were divided into an Experimental and a Control group on the basis of: (a) age, (b) raw scores on the *Otis Self Administering Tests of Mental Ability, Higher Examination: Form B for High Schools and Colleges;* (c) initial reading rate as determined from *Silent Reading Check RRRP G8-5;* (d) comprehension scores on above test; and (e) the mean educational level in terms of years of school completed. Table 1 shows the mean scores for each group.

Table 1

Comparison of Experimental and Control Groups

| Variables | Experimental Group Mean | Control Group Mean |
|---|---|---|
| Age | 25.1 | 25.0 |
| Otis Test | 56.2 | 56.5 |
| Rate (words per minute) | 351.8 | 352.4 |
| Comprehension (per cent) | 84.1 | 83.5 |
| Education in Years | 14.7 | 14.9 |

In addition to these factors, others such as the number of regular and reserve officers, years of service, active and inactive, were also considered and equated.

## Program of Instruction

*Total Group.* The group received an hour of orientation which included a statement of objectives and procedures as well as familiarization with the various routine functions. There were two hours devoted to testing pre-

[1] The author wishes to express his gratitude to Colonel C. W. Shelburne, Director, and Lt. Col. E. B. Watson, Director of Instruction, Marine Corps Supply Schools, whose interest and efforts made this project possible. In conformance with existing Naval Regulations, the opinions or assertions contained in this article are the private ones of the writer and are not to be construed as official or reflecting views of the Navy Department or naval service at large.

viously described; i.e., visual, reading, and intelligence. At the conclusion of the course, an hour of critique was held for a final summary of progress.

*Reading Rate Controller.* This device was used by both the Experimental and Control groups. Each group spent eighteen 25-minute training sessions on the Reading Rate Controller.

*Control Group.* This group followed the normal course of instruction which included the Reading Rate Controller training described above. In addition, there were eighteen 12½-minute training sessions with the tachistoscope. This training provided practice at 5, 6, 7, 8, and 9 digits at 1/25 and 1/100 of a second. Progression to the next higher series of digits was controlled by successful completion at each speed setting. Successful completion was interpreted to be 23 correct from a group of 25 slides. The limits for competency were those adopted during the research phase of the reading program sponsored by the U.S. Navy Field Medical Research Laboratory. In addition to the instruction and practice on the devices, there were nine 12½-minute periods devoted to the development of vocabulary and an equal number for the development of comprehension skills. During the vocabulary periods, there was instruction presented in independent word attack skills and dictionary skills. Vocabulary checks were administered at each session which served to introduce new skills as well as to provide practice. In the alternate sessions, emphasis was placed upon the development of comprehension skills. The student was provided the opportunity to practice in a free reading situation those skills being forced by the controlled reading training periods. Comprehension exercises were provided and served as the check valve for the tendency to overemphasize speed in reading.

*Experimental Group.* The training received by this group differed in that no tachistoscope was provided. The time normally allocated to this training was included in a broader program of training in vocabulary and comprehension skills.

## Results

The results presented are those dealing with the differences that existed between the Experimental and Control groups in the area previously discussed. The results with the exception of reading rate shall be based upon the quantitative measurements made from the Ophthalmograph film records.

Table 2

Summary of Mean Reading Rates in Words per Minute

| | Initial | Final | Diff. | $s_{x_1-x_2}$ | $t$ |
|---|---|---|---|---|---|
| Experimental | 351.8 | 809.8 | 458.0 | 40.5 | 11.314* |
| Control | 352.4 | 571.8 | 219.4 | 34.1 | 6.435* |
| Diff. | 0.6 | 238.0 | | | |
| $s_{x_1-x_2}$ | 5.8 | 56.9 | | | |
| $t$ | 0.1 | 4.184* | | | |

* Significant at the 1% level.

*Reading Rate.* The changes that occurred during the training period are shown in Table 2.

The data fail to indicate any significant difference between the groups prior to instruction. At the end of the training period, a statistically significant difference between the groups does appear. It will also be noted that both groups made significant gains in reading rate as is indicated by their initial and final reading levels. In considering the total group, the Initial Reading Rate was 352.1 and the Final Reading Rate was 690.8. The difference of 338.7 and the standard error of 33.3 resulted in a "$t$" of 10.165 which is significant at the 1 per cent level.

*Number of Fixations.* In Table 3 there is a comparison made of the mean number of fixations for each group during the reading of a 100-word passage.

There are no significant differences between the groups indicated either prior to or following the training program. There are statistically significant differences found within

Table 3

Summary of Mean Number of Fixations for 100-Word Passage

| | Initial | Final | Diff. | $s_{x_1-x_2}$ | $t$ |
|---|---|---|---|---|---|
| Experimental | 77.5 | 52.1 | 25.4 | 3.6 | 6.987* |
| Control | 76.4 | 54.8 | 21.5 | 4.8 | 4.476* |
| Diff. | 1.2 | 2.7 | | | |
| $s_{x_1-x_2}$ | 4.9 | 3.7 | | | |
| $t$ | 0.238 | 0.740 | | | |

* Significant at the 1% level.

Table 4

Summary of Mean Number of Regressions for 100-Word Passage

| | Initial | Final | Diff. | $s_{D_1-x_2}$ | $t$ |
|---|---|---|---|---|---|
| Experimental | 11.8 | 5.2 | 6.6 | 1.9 | 3.439* |
| Control | 8.3 | 4.1 | 4.2 | 2.8 | 1.515 |
| Diff. | 3.4 | 1.1 | | | |
| $s_{D_1-x_2}$ | 3.1 | 1.5 | | | |
| $t$ | 1.116 | 0.722 | | | |

* Significant at 1% level.

each group between the initial and final means. As a total group, the mean number of fixations were reduced from 76.9 to 53.5. The "*t*" value of 7.855 is significant at the 1 per cent level.

*Number of Regressions.* Since the reduction of the number of regressions would be another expected outcome of visual training, the mean number of regressions was computed for each group from Ophthalmograph records. These are shown in Table 4.

There are no indications of significant differences with the exception of those found within the Experimental group. As a total group, the mean was reduced from 10.1 to 4.7. The "*t*" value for the difference was 3.220 which is significant at the 1 per cent level.

*Average Span of Recognition.* This measure is directly related to the number of fixations since it is computed from them. It is presented as additional information in Table 5.

There are no indications of significant differences between the groups either prior to or following instruction. There are significant differences noted within each group. As a total group, the mean was increased from 1.37 words to 1.97 words. The "*t*" value of 9.071 is significant at the 1 per cent level.

*Average Duration of Fixations.* This measure is derived from the number of fixations and the reading time for the 100-word passage. The data are shown in Table 6.

Table 5

Summary of Mean Spans of Recognition in Number of Words

| | Initial | Final | Diff. | $s_{D_1-x_2}$ | $t$ |
|---|---|---|---|---|---|
| Experimental | 1.37 | 2.04 | 0.68 | 0.09 | 7.424* |
| Control | 1.37 | 1.90 | 0.53 | 0.10 | 5.503* |
| Diff. | 0.01 | 0.14 | | | |
| $s_{D_1-x_2}$ | 0.08 | 0.13 | | | |
| $t$ | 0.069 | 1.085 | | | |

* Significant at the 1% level.

Table 6

Summary of Average Duration of Fixations in Seconds

| | Initial | Final | Diff. | $s_{D_1-x_2}$ | $t$ |
|---|---|---|---|---|---|
| Experimental | 0.20 | 0.19 | 0.01 | 0.01 | 0.737 |
| Control | 0.19 | 0.20 | 0.01 | 0.01 | 1.594 |
| Diff. | 0.01 | 0.01 | | | |
| $s_{D_1-x_2}$ | 0.01 | 0.01 | | | |
| $t$ | 1.539 | 0.641 | | | |

The data present no basis for assuming any differences between the groups or within each group as a result of training. The total group initially had a mean of 0.19 second, and a final of 0.20. The difference was not significant, but the means indicate that there is agreement with the general acceptance of the average duration of fixations.

## Summary

In comparing the results of the Experimental and Control groups, there were no significant differences between the groups in the reduction of the number of fixations, the increase of the span of recognition, the reduction of regressive movements, or reduction of the duration of fixations. There was a significant difference in reading rate at the conclusion of the training program, but this was in favor of the Experimental group. Significant increases were observed within the groups upon comparing the initial and final performances in the area of number of fixations, span of recognition, and reading rate.

1. In the adult population studied, the results failed to indicate that the Experimental group was penalized through a lack of tachistoscopic training.

2. Although the areas measured were those in which tachistoscopic training is expected to produce positive results, the data failed to provide any basis for rejecting the null hypothesis.

3. Improvement of the visual skills measured by the Ophthalmograph may be the *result* of the improvement in reading efficiency rather than contributing to this improvement in reading.

*Received January 25, 1952.*

# The Effect of Fluorescent Flicker on Visual Efficiency

**Attilio Zaccaria, Jr., and M. E. Bitterman**

*University of Texas*

Although fluorescent lighting has come into increasingly widespread use, there is still considerable doubt about its suitability. Reports of strain and discomfort from workers and students required to perform visual tasks under fluorescent lighting are numerous (**6, 12**), but the source of these difficulties has yet to be established. As Luckiesh (**8**) noted, it is necessary to distinguish between the quality of light and the quality of lighting. Luckiesh himself believes that complaints about fluorescent lighting may be traced to inadequate control of glare, an opinion in which Hardy (**4**) concurs, and the well-known resistance of the public to technological change may play some role. The possibility remains, however, that intrinsic properties of the new source—particularly spectral characteristics and flicker—may be responsible in part for the reported difficulties. Questions relating to ultraviolet output, lag of emission in the yellow-green regions, and the significance of spectral lines have been widely discussed, but no generally acceptable conclusions have been reached (**12, 10, 7, 5**). Flicker is generally recognized to be undesirable, but many writers have minimized the importance of this factor in 60-cycle installations (**12, 2**).

There has been surprisingly little direct study of visual functioning under fluorescent illumination. Luckiesh and Moss (**9**) found no significant difference in frequency of blinking during work under tungsten-filament and fluorescent daylight lamps and concluded that ease-of-seeing was the same under the two conditions. Unfortunately, however, the unreliability of the index employed and its lack of relation to effort-expenditure in visual work casts doubt upon the validity of this conclusion (**20**). Preference tests conducted by Holway and Jameson (**6**) gave a clear advantage to tungsten as compared with a variety of differently colored fluorescent sources, but the validity of preference tests cannot be taken at face value. Dark-adaptation was found to be retarded by work under fluorescent as compared with incandescent light, but the difference disappeared when indirect lighting was used. In a study with the American Optical Company's Sight-Screener, Gray and Prevetta (**3**) recorded changes in a variety of visual functions following two hours of reading under equally bright (20 foot-candle) fluorescent lighting and daylight. No significant difference between the two conditions was found, but the possibility remains that differences might have been revealed by other measures.

If research on visual functioning under fluorescent light has not been abundant, neither has it been sufficiently analytical; no attempt has been made to separate the two principal factors—spectral characteristics and flicker—which differentiate fluorescent from other sources. In the experiment here reported, which was designed to study the effects of flicker, that factor was isolated by comparing A.C.- and D.C.-operated fluorescent lamps.[1] Furthermore, there has been no attempt in previous investigations to insure that the amount of visual work done by the subjects under the conditions studied was the same, a difficulty which was remedied in the present study by the use of a visual task that provided an objective measure of performance. Finally, insufficient attention has been given to the selection of appropriate indices of fatigue and efficiency. This problem is the most crucial of all in contemporary research on illumination (**1**); performance measures are inadequate indices of fatigue and efficiency, and techniques for the evaluation of effort are still in a developmental stage (**13, 19**). In the present study, primary concern with the effects of fluorescent lamp flicker led to the use of CFF (critical fusion frequency) as an index of fatigue. There is some evidence that CFF is reduced by exposure to flickering light (**18, 11**), and a variety of studies by Simonson and

[1] In order to study the effects of spectral quality independently of flicker, all sources should be D.C.-operated.

Table 1

Design of Experiment

| Group | Test Form | | Illumination | |
|---|---|---|---|---|
| | Period 1 | Period 2 | Period 1 | Period 2 |
| 1 | I | II | A.C. | D.C. |
| 2 | I | II | D.C. | A.C. |
| 3 | II | I | A.C. | D.C. |
| 4 | II | I | D.C. | A.C. |

his colleagues (**14**, **15**, **16**, **17**) suggest that a decline in CFF may be employed as a valid index of impairment.

## Apparatus and Procedure

The visual tasks selected for the experiments were Forms I and II of Tinker's *Speed of Reading Test*.[2] This test consists of a series of numbered sentences, each of which contains a word which spoils the meaning of the sentence, and the task of the *S* is to locate and cross out these words. The sentences are simple enough to minimize the comprehension factor when the test is used with college students. The number of items correctly marked in a given period provides an objective measure of achievement.

Twenty students at the University of Texas, 12 males and 8 females, served in the experiment. Their vision was normal or corrected to normal. Each *S* took both forms of the test, 30 minutes being allowed for each form. Half the *S*s began with Form I and half with Form II. Each *S* worked under both conditions of illumination, half beginning with A.C. and half with D.C. The procedure, therefore, required four subgroups, as shown in Table 1.

The *S*s were studied individually in a specially constructed, semicircular enclosure, 6 feet high and 41 inches wide, which contained a writing surface. The entire surround was painted flat white. Three 15-watt General Electric daylight fluorescent lamps were mounted on the ceiling of the booth, well out of *S*'s visual field. The power-supply made it possible to operate the lamps, which were wired in series, on either A.C. or D.C. The transition from alternating to direct current, and the reverse, were accomplished by the operation of a double-pole-single-throw switch. The level of illumination on the working surface was 20 foot-candles under each condition. A circular opening, 5/8 inches in diameter, was cut in the surround at eye-level. The opening, covered with a milk-glass plate, was 15 inches away from *S*, and behind it a General Radio Strobotac, Type 613-B, was set. With this apparatus the CFF determinations were made.

*S* was uninstructed as to the purpose of the experiment. He was told that his speed of reading was being measured and he was asked to work as rapidly and as accurately as possible. For each *S*, fifteen CFF determinations (ascending order) were made in all—five before the reading began, five at the conclusion of the first reading period, and five at at the end of the second period.

## Results and Discussion

Table 2 shows the mean performance scores and CFF drops under the two conditions of illumination. The difference in performance was not statistically significant ($t = .7$, $P > .05$). The drop in CFF under the D.C. condition was not significant ($t = .75$, $P > .05$), but the A.C. drop was hightly significant ($t = 5.0$, $P < .01$). The difference between the two conditions in terms of CFF drop also was statistically significant ($t = 2.9$, $P < .01$). Changes in CFF for Groups 1 and 3 and Groups 2 and 4 combined are plotted in Fig.1.

A more thorough examination of the relationships revealed in the data was made in an analysis of variance. The principal purpose of this analysis was to test the signifi-

Table 2

Speed of Reading and CFF Drop as a Function of Illumination

| | A.C. | D.C. | Diff. | S.E. Diff. | N | *t* |
|---|---|---|---|---|---|---|
| Reading Score | 272.8 | 270.4 | 2.4 | 3.4 | 20 | .7 |
| CFF Drop | 1.25 | | | .25 | 20 | 5.00 |
| | | .51 | | .68 | 20 | .75 |
| | | | .74 | .25 | 20 | 2.96 |

[2] We are indebted to Dr. Miles A. Tinker for providing these materials.

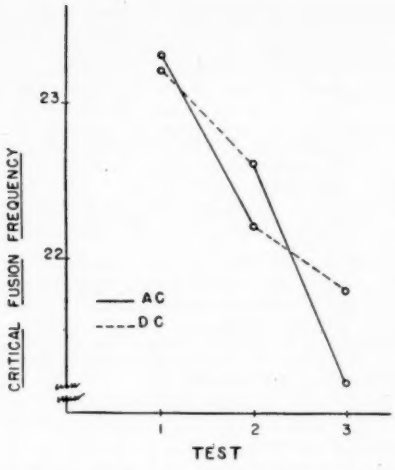



FIG. 1. CFF drop as a function of illumination.

cance of the order effect suggested in the curves of Fig.1, but this effect did not prove to be significant ($F = 1.1$, $P > .05$). Analysis of variance confirmed the conclusions based on t-tests. CFF drop was significantly influenced by condition of illumination, but speed of reading was not.

At the conclusion of the experimental session, each *S* was asked whether he had noticed a difference in the illumination employed in the two periods of work. Of the 20 *S*s, only five noted a difference, but all five, without exception, reported a preference for D.C. condition, which came first for two of them and second for the remaining three. These reports, therefore, tend to confirm the objective findings of the experiment.

These results suggest the undesirability of single-lamp A.C.-operated fluorescent fixtures which are widely used in the home, if not in commercial installations. The present experiment permits no conclusion with respect to the adequacy of two-lamp, out-of-phase installations such as are commonly used in factories and schools, but these fixtures could be studied by the technique here employed. This experiment provides no basis for evaluating the spectral properties of fluorescent sources, although again it suggests the kind of work which might be done. In order to control the flicker factor, D.C.-operated fluorescent lamps should be used for comparisons with other sources.

## Summary and Conclusions

The experiment here reported was designed to study the effects of fluorescent lamp flicker upon visual fatigue. Performance of a standardized visual task was measured for two 30-minute periods under 20 foot-candles of fluorescent daylight illumination. During one of the periods the lamps were operated with direct current, while during the other period they were operated with alternating current. In this way spectral characteristics, brightness level, and distribution were constant, with flicker being the only variable. Performance did not differ significantly under the two conditions, but the A.C. condition produced a significantly greater drop in critical fusion frequency than did the D.C. Only 25 per cent of the subjects detected a difference between the two conditions of illumination, but those that did uniformly expressed a preference for the D.C. condition.

It may be concluded that single-lamp, or in-phase, multiple-lamp fluorescent installations are undersirable. Further experimentation is needed for the evaluation of out-of-phase installations.

*Received December 24, 1951.*

## References

1. Bitterman, M. E. Lighting and visual efficiency: The present status of research. *Ill. Engng.*, 1948, **43**, 906–922.
2. Council on Industrial Health. The effect of fluorescent lighting on vision. *J. Amer. Med. Assoc.*, 1945, **128**, 1229.
3. Gray, J. S., and Prevetta, P. Fluorescent light versus daylight. *J. appl. Psychol.*, 1950, **34**, 235–236.
4. Hardy, L. H. Discussion. *Ill. Engng.*, 1945, **40**, 289.
5. Harmon, D. B. Lighting and the eye. *Ill. Engng.*, 1944, **39**, 481–500.
6. Holway, A. H., and Jameson, Dorothea. *Good lighting for people at work in reading rooms and offices.* Graduate School of Business Administration, Harvard University, 1947, 1–43.
7. Ley, E. B. Study of illumination. *Ill. Engng.*, 1944, **39**, 501–505.
8. Luckiesh, M. Discussion. *Ill. Engng.*, 1945, **40**, 285.

9. Luckiesh, M., and Moss, F. K. Vision and seeing under light from fluorescent lamps. *Ill. Engng.*, 1942, **37**, 81–88.
10. Luckiesh, M., and Taylor, A. H. Radiant energy from fluorescent lamps. *Ill. Engng.*, 1945, **40**, 77–88.
11. McFarland, R. A., Holway, A. H., and Hurvich, L. M. *Studies in visual fatigue.* Graduate School of Business Administration, Harvard University, 1942, 191–219.
12. Morgan, L. D. There is something wrong with our fluorescent lighting applications. *Ill. Engng.*, 1945, **40**, 275–285.
13. Ryan, T. A., Cottrell, C. L., and Bitterman, M. E. Muscular tension as an index of effort: The effect of glare and other disturbances in visual work. *Amer. J. Psychol.*, 1950, **53**, 317–341.
14. Simonson, E., and Brozek, J. A note on methodological evaluation of selected visual tests. *Amer. J. Ophthal.*, 1948, **31**, 979–984.
15. Simonson, E., and Brozek, J. Effects of illumination level on visual performance and fatigue. *J. Opt. Soc. Amer.*, 1948, **38**, 384–397.
16. Simonson, E., and Brozek, J. The effect of spectral quality of light on visual performance and fatigue. *J. Opt. Soc. Amer.*, 1948, **38**, 839–840.
17. Simonson, E., and Enzer, N. Measurement of fusion frequency of flicker as a test of fatigue of the central nervous system. *J. Indus. Hyg. & Toxicol.*, 1941, **23**, 83–89.
18. Snell, P. A. An introduction to the study of visual fatigue. *J. Soc. Motion Pic. Eng.*, 1933, **20**, 367–390.
19. Travis, R. C., Kennedy, J. L., Mead, L. C., and Allphin, W. Muscle action potentials as a measure of visual performance cost. *Ill. Engng.*, 1951, **46**, 182–187.
20. Wood, C. L., and Bitterman, M. E. Blinking as a measure of effort in visual work. *Amer. J. Psychol.*, 1950, **63**, 584–588.

# Visual Tracking: II. Effects of Brightness and Width of Target [1]

**Robert S. Lincoln and Karl U. Smith**

*University of Wisconsin*

The problem of the role of sensory cues in defining accuracy in visual tracking has both theoretical and applied significance. Elementary notions suggest that precision in tracking would correspond directly to the perceptual cue refinement resulting from size, brightness and velocity variations in the sensory situation. But visual tracking is not a simple, direct response to sensory cues. Rather, it is a response pattern composed of simultaneous adjustive reactions of position, rate, and acceleration control. Furthermore, the human operator in the tracking task acts as an oscillating generator, whose accuracy of response is only partly defined by the perceptual aspects of the tracking situation. The tracker's precision, therefore, may not be directly related to the magnitudes of stimulation produced by the sensory characteristics of the target and of the cursor.

In regard to these special considerations about visual tracking, an attempt has been made to determine in fact how tracking accuracy varies with brightness of target area, width of target, and with changes in the pattern of the target and cursor.

## Experimental Methods

1. *Design.* In this report we describe a series of experiments, all dealing primarily with the effect of changes in target width upon the accuracy of tracking. Three experiments have been performed in order to determine the nature of this target-width effect relative to variation in target brightness and target-cursor pattern.

The first experiment utilized a design in which four subject groups were trained and tested on different target-cursor patterns combined with targets of different width. The subject groups were all tested over the same range of variation in target brightness.

In the second experiment, the effect of variation in target width upon the accuracy of tracking was studied with a target and cursor of fixed pattern and fixed brightness. This experiment was conducted in order to determine the character of target-width effects with a greater variation in the width of the target. The study involved a Latin square design in which all subjects were tested under all experimental conditions.

The third experiment dealt with target-width effects in tracking when cursors of different colors were used in combination with a number of target widths. This experiment also utilized a Latin square type of design in which all subjects were tested under all experimental conditions.

The specific conditions of these experiments and other pertinent details connected with the statistical aspects of design will be mentioned in connection with the presentation of the results.

2. *Apparatus.* The detailed features of the apparatus used in these experiments have been described previously.[2] The subject's task is to track a small target rotating at varying speeds. The target moves in both a clockwise and counterclockwise direction through a limiting arc of some 300 degrees. Its radius of rotation is approximately nine inches. A cam arrangement regulates the direction and velocity of target movement. The subject controls a handwheel, 18.4 cm. in diameter, located approximately 130 cm. from the target plane. This handwheel in turn controls a cursor of the same radial length as the radius of rotation of the target. Cursor and target approximate the same plane of movement in order that parallax effects may be eliminated.

Target-cursor pattern is varied by sliding the cursor up or down so that either an overlapping or vernier pattern can be secured. Variations in target width are effected by means of a slide arrangement on the target support.

The error-recording system utilized in the present apparatus produces an integrated error record that is registered on a clock. Such a system weights the subject's errors according to the magnitude of the error and, therefore, includes more information in the final trial score than would a system which records only "on target" time. The error recording adjustment of the apparatus was kept constant for all target widths. Accordingly, the intrinsic

[1] This research was supported by funds provided by the Legislature of the State of Wisconsin and assigned by the Research Committee of the Graduate School, University of Wisconsin.

[2] Lincoln, R. S., and Smith, K. U. Transfer of training in tracking performance at different target speeds. *J. Appl. Psychol.*, 1951, **35**, 358-362.

difficulty of the tracking task remained at a fixed level, regardless of target width.

The mechanism of the integrated error recording system consists of a hollow brass tube that is rotated at a constant speed by a motor. A triangular piece is cut from the surface of the tube and the tube is filled with sealing wax. A mechanical indicator makes continuous contact with the surface of the tube and the intermittent electrical contact between the indicator and the rotating brass plate of the tube registers magnitude of error in accordance with the deviations of the indicator from a zero position. This time of contact between the indicator and and the tube plate is registered on the clock.

Illumination of the target area is provided by a slide projector that is mounted on a tripod and enclosed except for the emerging light beam. Variation in the brightness of the target area is produced by a series of Wratten filters.[3]

## Results

*Experiment 1. Effects of Variation in Target Width, Target-Cursor Pattern, and Target Brightness.* In this experiment two target widths were used. The width of one target subtended a visual angle of 42 minutes while the visual angle subtended by the second measured only 4 minutes. The targets were 60 minutes high. The cursor used was 4 minutes wide and 60 minutes long. Target-cursor pattern was varied in two ways. In one pattern the cursor completely overlapped the target, while, in the second, the cursor and target were arranged in a vernier relationship. Both target widths were used with both types of cursor, making up four separate target-cursor combinations. Target brightness was varied over four levels which covered a range of three log units.

Twenty-four students, both male and female, served as subjects in this experiment. The instructions to the subject were: "Try to keep the cursor in the center of the target at all times." These instructions were constant in all experiments.

On the first day of the experiment all subjects received four two-minute trials. Each trial involved a different target-cursor combination. The orders in which the different combinations were presented to the 24 subjects were assigned at random from the total of 24 possible orders. On the second day of the experiment all subjects again received four two-minute trials involving the four target-cursor combinations, but in a randomly assigned order which differed from that of the previous day. The mean scores obtained on these first eight trials served as a basis for matching the subjects in four groups of six subjects each.

On each of the remaining experimental days all of the groups tracked for three two-minute trials per day under one of the levels of target brightness. At the end of the experiment, all four groups had tracked under all four levels of brightness. The orders in which the subjects received the different brightness levels were also assigned at random so that each subject received a different one of the 24 possible orders. Six minutes of dark adaptation were required of all subjects before the first trial was begun on each of the last four days. Following the completion of this phase of the experiment, kymograph records were obtained of the adjustments made by some of the subjects in tracking the narrow and wide targets.

The accuracy scores obtained indicated that the wide target was superior to the narrow target when all other conditions were considered together. Generally, the vernier cursor was superior to the overlapping cursor. Finally, tracking accuracy was observed to be

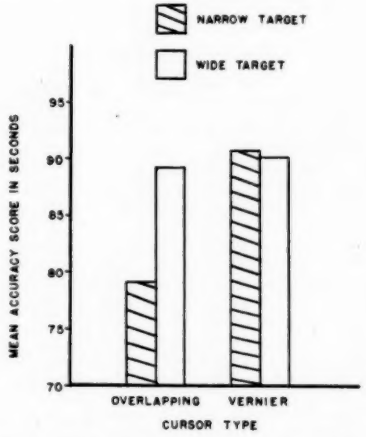



FIG. 1. Mean accuracy scores made with different cursor types and with different target widths.

[3] Eastman Kodak Company, Rochester, New York.

directly related to the level of target brightness. An analysis of variance performed on these data showed differences between the levels of these variables to be significant below the .1% point.[4]

Additional information concerning the relationship between target width and tracking accuracy is pictured in Figure 1. This figure shows that the accuracy of tracking the wide target was greater than the accuracy obtained with the narrow target only when the cursor was in the overlapping position. When the cursor was in the vernier position, little difference appeared between the targets. This interaction was significant below the .1% point.

Further evidence of the complicated nature of the relationships existing between the variables studied in this experiment is shown in Figure 2. The interaction represented by this figure was also significant below the .1% point. The figure portrays the same relationship depicted in Figure 1, but in addition indicates that even these relationships vary with the level of target brightness. At the higher brightness levels the wide-target overlapping-indicator combination seems to be superior to all others. However, "*t*" tests show that this target indicator combination is significantly superior to the narrow-target vernier-indicator combination only at the brightness level of log 0.37 millilamberts. At the lower brightness levels, the vernier indicator was superior regardless of target width. As the figure also indicates, the lowest accuracy scores were made throughout the entire range of target brightness when a narrow target was combined with an overlapping indicator. Similar results appear in an examination of the kymograph records obtained at the end of this experiment.

*Experiment 2. The Effects of a Wide Range of Target Widths, and Variation in Cursor Color.* In this experiment six target widths were used. The targets were white and subtended visual angles of 4, 30, 56, 82, 108, and 134 minutes. Their height subtended a visual angle of 58 minutes. The cursor, which was

[4] The critical summaries of data as well as the analysis of variance tables have been deposited with the American Documentation Institute.

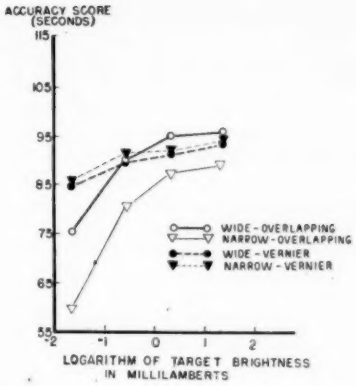



FIG. 2. The variation in tracking accuracy as a function of level of illumination of the target-cursor pattern for an overlapping pattern (open figures) and for a vernier pattern (solid figures) and with different widths of target.

kept in the overlapping position, had an angular width of 4 minutes and a length of 60 minutes. In this experiment, it was painted a flat red. A high level of target brightness was maintained throughout the experiment.

Subjects were assigned to the rows of three randomly drawn 6 × 6 Latin squares, in which the columns represented trials and the treatments were the six target widths. The 18 subjects were students who appeared on two separate days with a one-day interval between appearances. On the first day, all subjects tracked each of the targets for two minutes. On the second day, this pattern was repeated. At the end of the second day's trials all subjects ran one trial on the 4-minute target and one trial on the 30-minute target, this time using a white cursor. The order in which these targets were tracked was counterbalanced between subjects.

Analysis of the difference in scores made with the 4- and 30-minute targets combined with the white cursor again showed that accuracy was greater with the wide target than with the narrow target. A "*t*" test indicated that this difference was significant below the 5% level. Relative to the mean score made with the narrow target, this difference represents a 2.8 per cent increase in accuracy.

The scores made with the red cursor showed no increase in accuracy for the 30-minute target as compared to the 4-minute target. Actually, the mean scores made in this situation were identical to the nearest one-tenth second for these two target widths. Of equal interest is the fact that over a range of target widths in which the widest target was over 30 times as wide as the narrowest, the accuracy of tracking decreased only 3.5 per cent. This, too, is a significant difference, but it is one of a surprisingly small magnitude considering the extent of the variation in target width.

*Experiment 3. Retest of the Effects of Target Width and Variation of Cursor Color.* In the second experiment, the color of the cursor was observed to have some effect on accuracy when only two target widths were used with both cursor colors. The third experiment was conducted to test the color effect over a greater range of target widths.

Five targets having angular widths of 6, 18, 30, 42, and 54 minutes were used. These targets were painted a flat white and were 58 minutes high. Two overlapping cursors were used with these targets. Both cursors subtended visual angles of 6 minutes in width and were 66 minutes long. One was painted a flat white, while the second was painted a flat red. In the experiment, each cursor was combined with each of the five targets, making a total of 10 experimental conditions. Target brightness was maintained at a high level throughout the experiment.

As subjects, twenty students were assigned to the rows of two randomly drawn $10 \times 10$ Latin squares. All subjects appeared for five consecutive days, and on each day tracked for one minute with each of the 10 target-cursor combinations in an order indicated by the Latin square design. Scores from only the last three days were used in the determination of the results, because it was desired to test the critical effects on well-trained subjects.

The results with the red cursor confirmed Experiment 2 in showing that no optimum target width appeared. For this more restricted range of target widths, in which the ratio of the widths of the narrowest and widest target was 1/9, an accuracy decrease of only 0.9 per cent was observed.

With the white cursor there appeared a relative decrease in accuracy scores of 0.4 per cent when the widest target was compared to the narrowest. In addition, slightly higher scores were made with targets of angular widths, of 18, 30 and 42 minutes than with the narrowest target, although these differences did not meet the 5% criterion of significance.

When all targets were considered together, there appeared to be little difference between the mean scores made with the red and white cursors.

## Summary

The present study, comprising three experiments, has disclosed certain significant facts about the psychophysical determination of error in direct tracking of a target.

It has been found that the pattern relations of target and cursor are related to accuracy of visual tracking. A vernier relation of target and cursor generally produces more accurate performance than an overlapping pattern of target and cursor when the color of the two elements of the visual presentation is the same.

Illumination level of target and cursor affects accuracy in tracking. At low illuminations, accuracy drops off decisively for a cursor pattern which overlaps the target. Generally, the effects of change in illumination in relation to variations in the pattern of target and cursor, as well as alterations in the width of the target, are marked by complex interactions which make it difficult to formulate simple rules regarding these relations.

By far the most significant result of this study is the observation that an increase in the width of the target produces no marked change in the level of tracking accuracy. Targets some thirty times the width of the controlled cursor are tracked with almost the same accuracy as targets equal to the width of the cursor. In fact, under some conditions, maximum levels of accuracy are found with targets of a width greater than the width of the cursor. These results are secured under conditions in which the criterion of error remains fixed for all target widths.

The findings just noted have direct application for the design of the perceptual presentation in tracking devices as well as theoretical

implications for understanding tracking behavior.

The target-width effect, as observed here, suggests that the psychophysical laws of tracking behavior deal mainly with phenomena of organization of visual pattern rather than tolerances of alignment and misalignment of visual contours. The tracker's accuracy is determined by precision in scaling or bisection of the visual pattern rather than in terms of the optical resolution of limiting contours that define the visual presentation.

It may be that the target-width effect is a direct product of the character of the tracking response. The continued oscillation of the tracker about the target restricts visual judgment to a sustained estimate of relative symmetry of target and cursor position. In this sense, the present study constitutes a demonstration that the psychophysical aspects of tracking behavior reflect the special characteristics of variable error control that identifies the nature of the tracker's response.

This interpretation of the psychophysical organization of tracking implies that any cue aspect of the tracking situation which does not alter the critical oscillatory features of response will not change materially the level of accuracy in the situation. Such a systematic notion of tracking seems to give a reasonable account of some of the other special aspects of tracking, especially the phenomena of learning, the effects of change in speed of target, and the effects of transfer of training in relation to target speed.[2]

*Received January 8, 1952.*

# Comment on Darley's "Special Review"

**Kenneth Eells**

*U. S. Naval Personnel Research Unit, San Diego, California**

In the April, 1952, issue of *J. appl. Psychol.*, appeared a "Special Review" of a book, *Intelligence and Cultural Differences* of which I was a major contributing author. It is not clear exactly why this particular book was singled out for this flattering "special" rather than routine reviewing. A search of the back files of the *Journal* reveals only two other such instances in the past five years.

Since the book deals with a large-scale research study in an important (and controversial) area, three possibilities immediately come to mind: (1) The book possesses unusual merit and therefore is worthy of a special effort to bring it to the attention of the *Journal* readers; (2) the book is so extremely bad that it should be pointed out as a horrible example; or (3) the book deals with a controversial area and therefore will be of special interest to the *Journal* readers.

The first of these possibilities can be dismissed quickly, in view of the strongly derogatory nature of the review. While the writer is not in a position to pass an objective judgment with respect to the second possibility, it does seem odd, if this is what prompted the special attention, that the Program Committee of Division 5 of the APA selected this particular book, and the research which it reports, as the subject for a symposium at the recent Washington meeting. If the third reason is the correct one, the Editor of the *Journal* is presumably at fault in having a controversial book reviewed by an avowed opponent of the point of view expressed in the book without making provision for presentation of other points of view.

It might be, of course, that the pages of an APA journal were simply used to present a one-sided review of a controversial book and that the review was played up by the Editor as a "Special Review" to add prestige to the attack contained in the review. But one would come reluctantly to any such conclusion. APA journals have had, for the most part, a long and honorable history of fair play and scientific objectivity of which they can be proud.

Darley charges the authors of the book with lack of dispassionateness and objectivity, special pleading, unwarranted assumptions, confusion of hypotheses with conclusions, unsatisfactory sampling procedures, faulty interpretation of correlations, improper selection of items for study, a research design so planned as to maximize the chances of proving the hypotheses, inconclusiveness of findings, and failure to perform certain possible analyses of the data. *Not a single significant favorable comment is found in the entire review*, although the book appeared under the joint authorship of five individuals connected with reputable institutions and was published by a reputable university press.

Probably the most serious charges which Darley makes, because they attack the scientific integrity of the authors, are those dealing with confusion of hypotheses with conclusions, unwarranted assumptions, and a research design so planned as to maximize the chances of proving the hypotheses. One would expect anyone making such charges to be scrupulously careful to be accurate and to document such statements in detail. Darley fails in both respects.

Space forbids an adequate discussion of these points but a few general observations can, and should, be made. The writer wishes first, however, to apologize for the too-general nature of the comments which follow, and for his failure to document his statements more fully or to develop them to the point where the reader could make an intelligent judgment about the matters himself. He prepared such a statement, taking up each of Darley's charges and citing quotations from the book in each case to show that Darley had either read carelessly or had misinterpreted what he

* The opinions expressed are solely those of the author and are in no way official; nor are they to be construed as representing those of the U. S. Naval Personnel Research Unit or Bureau of Personnel.

had read, had misunderstood the basic nature of the research, was being impossibly perfectionist in his expectations, and in one case had "slanted" his review by selective citation. All this was spelled out with quotations and page references to document it fully. The Editor ruled that this statement was too long.

An abbreviated form was then submitted, which would have required four journal pages. This too was refused publication, even when the writer offered to pay the regular page rate for adding new pages to the *Journal*, as is done with "early publication" materials. He was told "Either you will provide a reply within 1200 words (1½ pages) or you will have no space at all." The Editor's letter openly took sides with the reviewer.

Scientific understanding can sometimes be advanced through scholarly controversy, but only when such controversy is based upon either carefully documented evidence or carefully reasoned analysis. Neither is possible within the space limitations which apparently apply to such controversies in *J. appl. Psychol.* This being the case it is particularly important that reviewers give well-balanced and objective presentations of such controversies.

Darley's charge that two of the authors were lacking in objectivity and "transmute assumptions and hypotheses into foregone conclusions" reveals that he simply did not read, or did not understand, Chapter 1 of the book, in which the structure of the book was explicitly described. The chapters to which Darley refers were there clearly described as introductory analysis of the issues for the purpose of providing insight into the setting of the problem. Any interpretation of these chapters as conclusions from the research study *was explicitly warned against.* The wisdom of such a presentation may fairly be questioned, but to charge that the chapters do something which they do not do is something else again.

Space will permit comment on only one of the assumptions which Darley refers to: "Davis and Havighurst apparently assume . . . that all methodological questions in the study of stratification have been solved." This is utter nonsense. No one who has worked in this area could possibly think that all methodological questions have been solved. At no point in the book do any of the authors say, or imply, any such thing. In no other area of research is it customary to imply that no research should be conducted until all methodological problems have been completely solved. Would Darley like to see all research in the field of guidance deferred until perfect procedures are developed for measuring adjustment of individuals? To such extremes does a reviewer go when he substitutes zeal for destruction for objectivity.

To charge the authors of a research study with having so designed the study as to maximize the possibilities of proving their hypotheses is about as far as one can go in charging lack of personal and scientific integrity. Darley's description of what he based this upon reveals a complete misunderstanding of the purpose of the research study he was reviewing. Unfortunately, to discuss this important point in sufficient detail to make it clear to one who probably has neither the book nor the review before him would require more space than is available.

The writer will be glad to supply a mimeographed copy of his more extended statement to any interested reader.

*Received August 27, 1952.*
*Published out of turn by the editor.*

# Reply to Eells' Comment on Darley's "Special Review"

**John G. Darley**

*Laboratory for Research in Social Relations, University of Minnesota*

The editor of this journal has permitted me to make a short rejoinder to the above comment by Dr. Eells on my earlier review of the book of which he was senior author. I doubt that this exchange will produce any *rapprochement*, in view of the complexities and involvements inherent in research investigations of stratification phenomena, but I shall try at least to make my position as a reviewer clear.

A substantial part of Dr. Eells' comment concerns the editorial policies and practices of this *Journal*. I am obviously not in a position to reply to this, nor is it entirely germane to the issue as I see it. I am sure that Dr. Eells' position on my review will be abundantly clear to anyone who accepts his offer to receive "a mimeographed copy of his more extended statement." I sincerely hope many psychologists take advantage of that offer.

The burden of his present comment regarding my review appears to be twofold: (1) I made no significant favorable comment in my review; and (2) I have criticized him and particularly his co-authors not only for the design and interpretation of his study but also for their interpretations and treatment of the general problem toward which his research was directed as a specific study.

I know of no commonly accepted canons or mores of book reviewing for technical journals under which a reviewer is *required* to make favorable comments. I have, rather, operated on the philosophy that if a reviewer believes a book to be either outstandingly bad or good, he is morally obligated to state his carefully considered opinions unequivocally. This obligation is the greater, to the extent that a book deals with a serious or a controversial topic and to the extent that great claims are made for the book. Psychology should by now be somewhat beyond the point in its history as a science when it must award many A grades for effort only.

In the last analysis, neither the author nor the reviewer represents the final judgment in history. Both the author and the reviewer express their own best judgments, morally and scientifically; present and future readers are still free to accept the judgment of one or the other, or neither, as they see fit.

So far as the second phase of Dr. Eells' comment is concerned, he has well summarized my objections to his work, and I find no reason to change them on the basis of his rebuttal. While I have no wish to choose up sides in this matter, it was of interest to note that McNemar's recent and independent review of the same book[1] raised some of the same criticisms that seemed to me worth noting, and in addition raised a few other objections that, regrettably, escaped me.

[1] McNemar, Q. Review of *Intelligence and Cultural Differences*. *Psychol. Bull.*, 1952, **49**, 370–71.

## Book Reviews

Dale, Ernest. *Planning and developing the company organization structure.* New York: American Management Association. Research Report Number 20. 1952. Pp. 232. $4.50. ($3.00 to AMA members.)

This book is the result of a two-year research program sponsored by the American Management Association and designed to emphasize the realities of organization. The research included visits to 40 companies of varying sizes; participation in the work of changing company organizations; examination and analysis of several hundred organization manuals and organization charts; a survey of American and foreign literature on organization; and interviews with many military and industrial thinkers and practitioners on organization.

There are two main parts to the book. Part One is entitled "The Dynamics of Organization—A Study of Organization Problems at Various Stages of Company Growth" and is divided into seven sections referred to as the seven stages of company growth—(1) determining objectives and dividing work; (2) delegating responsibility; (3) span of control; (4) the staff assistant; (5) the staff specialist; (6) group decision-making; and (7) decentralization. Part Two is entitled "The Mechanics of Organization" and is divided into 11 sections under such topics as reorganization; timing, defining and constructing the organization; the organization chart; the organization manual; and gaining acceptance. There are six appendices which include job descriptions of major executives; descriptions of management committees; content and extent of decentralization of major management functions; the summary of contents of an oil company management guide; and organization nomenclature of a large automobile company. In addition, there is an extensive bibliography grouped for each section of the book and also separate sections covering certain selected topics in organization.

Throughout all of the material the author combines what is technical, academic or scientific on the subject of organization with what he has found to be business practice and practical application. In all of the main sections and in many of the sub-sections he illustrates the theoretical material with pertinent cases from specific companies. In some instances, however, this process of illustrating leads him astray since he uses the specific case as illustrative of a general principle. Much of the material is illustrated with organization charts, although the principles and problems of designing and constructing organization charts are not given the importance which they justify in a book on this subject.

In his discussion of organization problems in Part One, the author stresses the desirability and the growing use of (1) the staff assistant, a person who is not an executive, but aids an executive in getting things done without taking away or relieving the executive of any of his responsibilities; and (2) the staff specialist or consultant. He emphasizes the value of the staff specialist in large organizations. However, he fails to point out the very pertinent fact that individuals who can be employed to perform specific assignments as the need for them arises are very valuable to small businesses where the quantity of work at no time justifies the employment of full-time persons for specific occasional duties. In a highly practical presentation he includes excellent discussions of group decision in operating organizations along with discussions of different types of committees, their functions, and the strategic factors in successful committee work.

Perhaps the major weakness of the volume is that throughout the many pages devoted to the concepts of responsibility and authority, the author does not define them, nor does he make any distinction between them. In fact, it appears that he uses the terms interchangeably, and discusses the *delegation* of responsibility as well as the delegation of authority. Since these concepts are important in organization planning, it would have been helpful if the discussions had emphasized that ideally the amount of responsibility and authority should be equal, that there are important differences between them, and that one of the outstanding differences is that *authority can be delegated whereas responsibility cannot.*

The professional value of this book to psychologists will be limited largely to those who have direct contacts with industry, or who are connected with schools of business, or who have a specific interest in executive activity and organization problems. While the author stresses the dynamics and continuity of organization planning and the dependence of much organizational activity on the personalities in the organization, his discussions do not include any direct references to the theoretical psychological principles involved. On the other hand, all business executives should become acquainted with the book. They should not read it, however, with the expectation that it will solve their organization problems or that it will eliminate the necessity of their continuing use of expert advice and counsel as the problems discussed in the book arise in their own business relationships.

C. G. Browne

*Wayne University*

Von Foerster, Heinz. (Ed.) *Cybernetics: circular causal and feedback mechanisms in biological and social systems.* New York: Josiah Macy, Jr. Foundation, 1952. Pp. 240. $4.00.

This is the third volume on cybernetics published by the Josiah Macy, Jr. Foundation and is a complete transcript of the eighth conference on this subject sponsored by the Foundation. Although the title does not indicate it, this conference was devoted almost entirely to problems of communication. Six papers were presented and thoroughly discussed by representatives of the natural and social sciences. Of particular interest to applied psychologists seriously concerned with communications might be the following: Communication Patterns in Problem Solving Groups (Bavelas), Communication between Men: Meaning of Language (Richards) and Communication between Sane and Insane: Hypnosis (Kubie). Communication between Animals (Birch), Presentation of a Maze-Solving Machine (Shannon) and In Search of Basic Symbols (MacKay) are somewhat far afield, irrelevant or overly technical for the reader without a strong general interest in communications per se.

The book is not a collection of answers to communication problems or even a report of experimentation directed toward that end. It is instead a stimulating source of hypotheses and suggestions and an interesting illustration of the application of widely differing approaches to a common problem. The discussions are intense and wide-ranging and at times seem to get out of hand. The free exchange of ideas, however, seems very valuable and is certainly worth the occasional digression.

Two appendices deal with the nomenclature of information theory and references.

James J. Jenkins

*University of Minnesota*

Janis, Irving L. *Air war and emotional stress.* New York: McGraw-Hill Book Company, Inc., 1951. Pp. 257. $5.00.

On November 3, 1944 President Roosevelt directed the War Department to investigate systematically the role of air power in war and peace. The "human factor" was an integral part of that investigation by the U. S. Strategic Bombing Survey in Germany and Japan; Britain was not included in the scope of the study.

After the war, the Air Force, desiring a more complete coverage of psychological aspects of air warfare, contracted with the Rand Corporation to review the literature on the subject, including British studies and the protocols of the U. S. Strategic Bombing Survey reports on Germany and Japan. Prof. Janis' book is the result. In Part I he dispels some popular illusions about reactions to A-bombs among the survivors at Hiroshima and Nagasaki. There was psychological unpreparedness for the explosion and resultant nonadaptive behavior as an *immediate* reaction. Though personal involvement was more widespread, emotional disturbances were not markedly more severe than acute anxiety and depression noted among British and German populations after severe raids of conventional bombing. Prof. Janis continues, "The amount of defeatism at Hiroshima and Nagasaki was less than in other Japanese cities . . . . The attitudes of the A-bombed population were found to resemble those of people in the lightly bombed and unbombed cities rather

than in the heavily bombed cities . . . " (p. 60) and "the dominant psychological effects resulting from A-bomb disaster generally did not differ in any unique way from those produced by other types of bombing disasters" (p. 71).

As to reactions to conventional bombing, "mental breakdown, panic and mass demoralization rarely materialized" (p. 153). Chronic mental disorders notably increased only among those having phobias to loud noises, darkness, fire, and danger. Psychosomatic disorders were not numerous. Individual and group adaptive mechanisms absorbed most shocks and emotional reactions to bombing.

On the other hand, widespread emotional stress resulted from deprivation caused by bombings. Inadequate food and sanitary facilities, breakdown of transportation and public utilities, lack of retaliation against the continuing air raids all caused bitter resentment against home authorities for failure to provide protection. Lowered morale became apparent as crime and delinquency rates, black-marketing, and other symptoms of social disorganization increased. Gradually apathy and defeatism developed into defection.

In conclusion, Prof. Janis warns that it can happen here, that realistic education to the human factor problems posed by the H-bomb, radiation poisoning and biological warfare is needed; that there must be a greater awareness of "group identification as a major factor underlying efficient performance in the face of danger" (p. 214).

This is more than a psychologically oriented air warfare review of literature and data. It is a provocative glimpse into the future. Psychologists particularly will find it challenging and worthy of study.

Milton D. Graham

*Human Resources Committee,*
*Research and Development Board,*
*Washington 25, D. C.*

Rogers, C. R. *Client-centered therapy.* Boston: Houghton Mifflin Co., 1951. Pp. vii + 560. $4.00.

This latest volume from the prolific Dr. Rogers is a systematic and inclusive description of nondirective therapy, as currently conceived. Part I presents a "current view," Part II, "the application," and a briefer Part III, "implications for psychological theory." In the "application" unit, Elaine Dorfman has contributed a chapter on "Play Therapy," Nicholas Hobbs, a chapter on "Group-centered psychotherapy," and Thomas Gordon, "group-centered leadership and administration." Gordon's chapter is distinguished by its extended and varied references, and by its willingness, at this stage of knowledge to accede flexibly to demands of practical situations.

In *Client-Centered Therapy* Carl Rogers gives further evidence of his profound faith in the ability of the individual to solve his own problems. His present "theoretical" position is that the well-adjusted individual is one whose self-structure can assimilate and include perceived experiences without threat. Many individuals, however, seek therapy because of the unsuccessful attempt of the self-structure to defend itself by rejection of those experiences which are inconsistent with the individual's conception of himself. If the counselor can indicate to the client unreserved acceptance and accurate understanding of the client's attitudes and perceptions, the client is permitted to become aware of and to integrate these previously "denied experiences" because they become gradually less threatening to the self.

Much of Rogers' message is buttressed by evidence from numerous researches of present and former students and by a wealth of illustrative excerpts which are taken verbatim from actual counseling interviews. Out of his conviction and an intimate, empirical acquaintance with therapy, he evolves 19 propositions, constituting the bulk of his last chapter, "a theory of personality and behavior."

For the personnel worker, the strength of *Client-Centered Therapy* probably lies in the coherent system which guides the interactions of counselor and client: the central "hypothesis" of nondirective therapy. Its fallibility appears to reside in the fact that many of the underlying constructs upon which the "hypothesis" is based seem to be intuitively rather than empirically derived. For example, "the phenomenal field," "the self" or "threat" cannot, to my knowledge, be directly observed,

but only inferred. One wonders whether Rogers' own basic distrust of empiricism in data-gathering and theory construction has impelled him to state (Proposition VII, p. 494) that "The best vantage point for understanding behavior is from the internal frame of reference of the individual himself." Perhaps it is this distrust that prevents Rogers from analyzing with meticulous care much of the research data which he presents in support of his arguments.

Personnel workers in general, especially those engaged in counseling activities, would do well to become familiar with the contents of this book. Just as Strong and his coworkers have contributed materially to our knowledge of interests and their measurement through a combination of singleness of purpose with extensive research and writing over a period of years, so Rogers and his students have added to our knowledge about counseling in the past decade. Careful reading of *Client-Centered Therapy* cannot fail to stimulate the personnel worker in the direction of asking himself questions about what he is doing, what he knows and does not know. In some ways Rogers is not unlike Paracelsus who strove mightily 450 years ago to answer questions about man in his universe with meager data and inadequate conceptions available to him. As in the case of Paracelsus, Rogers' latest book will probably be enthusiastically accepted by some, violently rejected by others, casually received by few.

Harold B. Pepinsky

*Occupational Opportunities Service,*
*The Ohio State University*

Bryan, Alice I., *The public librarian* (*With a section on the education of librarians* by Robert D. Leigh). New York: Columbia University Press, 1952. Pp. 474. $6.00.

This volume completes the series of studies of public libraries made by the Public Library Inquiry of the Social Science Research Council and directed by Robert D. Leigh. It consists of Parts 1–3 which contain Dr. Bryan's survey of public library personnel and library personnel organizations and administration; Part 4, which is a survey of the education of librarians (Dr. Bryan designed the survey and determined its method, though the execution was undertaken by Dr. Leigh); and, Part 5 which is the summary and conclusions for both surveys.

Dr. Bryan has done a workmanlike job in presenting and analyzing the information obtained from questionnaires and tests given to 3,706 librarians. In addition she has supplemented her data in some instances with other data. The data on personnel organization and administration were obtained by a fifty-six page questionnaire supplemented by personal visits to forty-three of the sixty libraries in the sample.

In analyzing the professional librarian Dr. Bryan studied personal characteristics (age, sex, marriage, personality revealed by the Guilford-Martin Inventory of Factors, recreation, etc.), Educational status (academic, professional, non-credit training, etc.), Economic status (salaries, promotion, personal savings, insurance, retirement, etc.) and the Library Career (vocational interests as measured by the Strong Vocational Interest Blank, present attitudes toward librarianship, attendance at professional meetings, professional reading, etc.). All in all this forms without question the most comprehensive picture of what the librarian is like, what he (or as Dr. Bryan aptly points out she) earns, and how well librarians are prepared to meet the library's objectives. It is unfortunate that this study appears in 1952, was written just a year ago and based on information obtained in 1947 and 1948. As a result some of the information presented, e.g., information on salaries, is not representative of the present time. And in interpreting the results the reader must bear this fact in mind although the author in several places points this out.

In the section on personnel administration Dr. Bryan presents detailed findings on how libraries handle their personnel problems. Chapters on personnel management, employee selection, classification, in-service training, employee welfare and morale, etc. present again the most comprehensive picture we have of library practices in the area of personnel administration. The conclusion that Dr. Bryan reaches is that personnel in public libraries is inadequately organized and supervised in terms of modern, scientific personnel manage-

ment. With this conclusion, the careful reader must agree for Dr. Bryan's data are complete and reliable. One feels at times however that she places too much emphasis on formal personnel procedures since many of the libraries surveyed were small ones (it appears that 20 of the libraries averaged less than 5 staff members apiece) where formal devices are less important than human, competent and discriminating chief librarians who use humanitarian principles in dealing with their few staff members. This is in contrast to the section on the librarian where in addition to Dr. Bryan's detailed facts one seems to receive some feeling of the character and personality of librarians.

Dr. Leigh's section on the education of the librarian gives information on the evaluation of library schools, their educational programs, the students and the faculty and instructional resources. The data were obtained by questionnaires to the thirty-four accredited library schools in the U. S.

Here again a complete picture is presented of library education in the U. S. In addition Dr. Leigh's study profits greatly from the fact that previous studies of library schools enabled the present data to be compared with the earlier figures and thus trends could be assessed. Dr. Leigh's study however gives more complete information than any of the others cited for earlier comparison.

Dr. Leigh's findings are in brief that there are inadequacies in the library schools, which are primarily the result of great disparities in resources among the schools themselves. He suggests that the way of improvement lies in the consolidation of facilities in the substantial number of strong library schools.

The two sections present interesting similarities and interesting differences. Both Dr. Bryan's section on personnel and Dr. Leigh's section on the education of librarians are well planned and well presented, the evidence collected is pertinent, it is summarized carefully and completely. In both instances we have the most complete picture available today. In both studies there are some interpretations that one might question but such instances are few.

The chief differences in the two sections are in respect to breadth and depth. Dr. Leigh's study seems to be broader in that it makes much greater use of previous studies both of library schools in general and of individual schools. Dr. Bryan's study is almost entirely limited to the presentation and analysis of the facts obtained. Even in analyzing the results she makes little use of other articles or books on library personnel administration.

To this reviewer at least Dr. Leigh's analyses and interpretations seem more penetrating. The discussion of library education seems to to have much more warmth, much more care and thought and more evidence of insight. Both however maintain the same high standards of the six volumes of the Public Library Inquiry previously published and the volume as a whole is a valuable addition to library literature.

Errett W. McDiarmid

*Dean, College of Science, Literature and the Arts, University of Minnesota*

Riley, John W., Ryan, Bryce F., and Lifshitz, Marcia. *The student looks at his teacher.* New Brunswick, New Jersey: Rutgers University Press, 1950. Pp. iv+166. $2.75.

This is a report of a single, but extensive, study made at Brooklyn College by the Sociology Department at Rutgers. Of the regularly enrolled Brooklyn College day students, 6,681 out of a possible 8,000 answered on one day an extensive questionnaire and deposited the results in a box anonymously. The book analyzes the returns.

The authors make a telling case *a priori* for the study of evaluation of teaching through student opinion. They point out that college teachers are well protected by tradition and rationalization from facing evaluation of any objective kind. They do not have the spur to improvement that is furnished by knowledge of success or failure. Teachers are selected for their training and research but not for success as a teacher. Too often promotion also is independent of teaching performance. In modern, large universities, the diagnostic effect of student election of courses is lost. Fellow faculty members do not visit each others' classes and depend largely on student gossip for their evaluations of each other.

It is, of course, entirely too much to expect that a single study, limited to its advance program, should supply the answers to the questions raised. Answers can come only as a result of long experience and of many approaches. Where shall we seek evaluations of teaching? From students, from colleagues, from past students, from administrators? How shall judgments be collected? To whom shall they be made available? What is their value?

The authors have assumed that the answer to the question of the best source of evaluative judgments is "the students." There are cogent arguments for this. Students are the only persons who observe teaching. They experience it as well as observe it. Teaching is a two-way operation and must be participated in by students or there is nothing taught.

The survey of which the book is a report and analysis made use of a questionnaire listing ten items important to good teaching. These were selected by the staff after reading the literature of evaluation and making their own additions and talking over the items with students. The ten traits were: (1) organization of subject matter; (2) speaking ability; (3) ability to explain; (4) encouragement to thinking; (5) attitude toward student; (6) knowledge of subject; (7) attitude toward subject; (8) fairness in examinations; (9) tolerance to disagreement; and (10) instructor as "human being." For each trait, four degrees were described. In the case of trait 5, for example, the extremes of the four degrees were described as: (a) sympathetic, helpful, actively concerned; and (b) distant, aloof, cold. By this device, the authors believe they were avoiding encouraging ratings on a comparative basis (Professor X as compared with Professor Z); they preferred to imply an absolute scale.

Students were asked possible relevant items, concerning the size of the class, their own grade point, whether the course taught was required or elective, and so on. Each student was asked to rate one of his instructors. The one chosen depended upon the last initial of the student and the serial order of the class in the weekly schedule. Findings in individual cases were reported in percentiles for each trait, based on the distribution (teachers of art, science, and social science taken separately).

The authors were not interested in the evaluation of the teaching of individuals, although each teacher was given the returns in his own case, indicating his percentile standing in each trait. No determinations of the reliability of these percentile scores were sought or reported.

The authors' main attention is given to the association of "traits" with such items as age, rank, student grade, student year in college, size of class, and field of study (art, natural science, or social science). Importance is attributed to the higher scores made by teachers on "knowledge of the subject" rather than on "encouragement to thinking," and to the influence from this fact that student judgments are in terms of an ideal norm.

Students' ratings of instructors are little affected by size of class, by the required or elective character of a course, or by the year in college. Value as a teacher appeared to diminish with the age of the teacher and with advancing rank. Value is favorably associated with the possession of a Ph.D. and by the publication of research.

The report will be valuable to anyone interested in the evaluation of teaching but it is disappointing in that the study was made almost without statistical analysis. The style of the book is reminiscent of federal agency reports, with the frequent use of such terms as "implement" to avoid being too precise.

E. R. Guthrie

*University of Washington*

Fuess, Claude M. *The College Board: Its first fifty years*. New York: Columbia University Press, 1950. Pp. vi + 222. $2.75.

The development of educational measurement in the United States has been an outgrowth of the insight and skill of many individual reasearch leaders, the combined effort of faculty members in a number of universities, the work of various governmental agencies including especially the Armed Forces, and the contributions of several institutions specializing in tests and examinations. Among Eastern colleges and preparatory schools, probably no institution has played a larger part in measurement theory and practice than the College Entrance Examination Board. In recent years,

this influence has been extended to other regions as well.

*The College Board* by Claude M. Fuess provides an intimate, detailed, and highly readable account of the first fifty years of the Board's organization, plans, and work. Perhaps no one was better qualified to write a book of this kind. As headmaster of Phillips Academy and as a leader in the independent school field, Dr. Fuess has been closely identified with the group of men and women who built the Board, a fact which he modestly overlooks in his book.

Dr. Fuess shows how the idea of a college board was born out of the need to correct the chaotic college entrance situation that prevailed during the latter part of the nineteenth century when each college prescribed its own entrance requirements without regard to what other colleges were doing. He then traces the development of the Board from its early organization down to the latest phase when it merged the operation of its testing activities with those of the Carnegie Foundation for the Advancement of Teaching and the American Council on Education to form the Educational Testing Service.

The book will be of more interest to educators than to psychologists. There are nine chapters in the book. The chapter on early organization is a particularly interesting and valuable contribution, not alone for its record of the formation of the Board, but also for the light it sheds upon two great and very different personalities in higher education who joined forces to establish the Board—Charles William Eliot and Nicholas Murray Butler.

Other chapters deal with the Board in action, the move toward comprehensive examinations, aptitude tests, problems and personalities, and the work of the Board during World War II and following the war. The book concludes with a brief chapter discussing the question of what lies ahead. In several chapters there is considerable detail concerning personalities well known in Board circles but perhaps of less interest to a wider reading public. But this kind of detail helps to create in the reader a feeling for the Board as an on-going organization.

Although there are occasional references to the work of measurement leaders outside the Board, such as Thorndike, Bingham, and Walter Dill Scott, the development of the Board's examinations is treated largely as a self-contained project. In point of time, the Board's history closely parallels the history of the measurement movement, and inevitably Board policy and the nature of its examinations were influenced by the activity and ferment taking place in educational measurement at all school levels throughout the country. However, *The College Board* was written for a special purpose, and it could hardly be expected that it would deal with the history of the whole field of educational measurement.

The style of Dr. Fuess's book, unlike that of many books in the field of education, makes for pleasurable, as well as profitable, reading. The book will be of interest and help to all who wish to have at hand a lively and accurate record of the first fifty years of an influential organization that has served institutions well and that may be expected to provide continued leadership and service of a higher order in the years to come.

Arthur E. Traxler
*Educational Records Bureau, New York City*

Hersey, Rexford. *Better foremanship—key to profitable management.* New York: Conover-Mast Publications, Inc., 1951. Pp. 244. $4.00.

This book is largely a collection of articles previously published by the author in *Mill and Factory, Personnel*, and *Personnel Journal*. It is intended for the foreman, ". . . give *him* the working tools of his profession, a basic conception of how to handle them, and a practical concept of his day-to-day job."

The book deals with the foreman's duties and responsibilities in the role of a leader, a practical psychologist, an administrator, a job analyst, an instructor, a safety engineer, and an executive. The last two chapters are directed to executives above the foreman level and deal with higher management's function in developing and maintaining better supervisors. There is a 56-page appendix consisting of a checklist for rating or self-analysis of foremen, and two questionnaires dealing with knowledge and

principles involved in supervision, complete with scoring keys.

The author does a complete job of covering in one volume most of the many and varied functions which make up the complex job of the present-day foreman. This complete coverage, however, results in one of the main weaknesses of the book. In order to cover so much material, the author has had to limit his discussion of a number of the subjects covered to a bare listing of rules-of-thumb for the foreman to follow. Many of these rules are stated flatly, with little or no explanation of the "why's" behind them. This is especially true in the case of the treatment of the "human relations" phase of the supervisor's job. This type of presentation creates the impression that a simple following of these rules-of-thumb is all it takes to become a successful and competent foreman. In fact, on page 14 we find the statement, "However, the Leader, who has followed the techniques laid down previously in this chapter need not worry about dissatisfaction in his group."

The book is written in a clear and concise style, with occasional examples and illustrations. Its use with foremen, however, will be limited somewhat because much of the vocabulary used is at too high a level to be understood by persons without a fairly extensive educational background.

Because the book contains so much diversified material—all of which should be included in a complete program of supervisory development—it should prove useful as a textbook supplement for the conference type of supervisor training. A discussion of the missing "why's" behind the many rules laid down in the book alone could well serve as the basis for a profitable series of training conferences.

Theodore R. Lindbom

*Personnel Department,*
*Midland Co-operative Wholesale,*
*Minneapolis, Minnesota*

# New Books, Monographs, and Pamphlets

Books, monographs, and pamphlets for listing and possible review should be sent to Donald G. Paterson, Editor, Department of Psychology, University of Minnesota, Minneapolis 14, Minnesota

*Vocabulary, semantics and intelligence.* Snowden Arthur. Bethesda, Md.: Lexicon Press, 1952. Pp. 115. $1.50.

*Social psychology.* Solomon Asch. New York: Prentice-Hall, Inc., 1952. Pp. 736.

*Readings in experimental industrial psychology.* Milton L. Blum, Editor. New York: Prentice-Hall, Inc., 1952. Pp. 480. $4.75.

*The selection of military manpower—a symposium.* Leonard Carmichael and Leonard C. Mead, Editors. Publication No. 209. Washington, D. C.: National Research Council, 1952. Pp. 270. $2.50.

*Infant and maternal care in New York City.* E. H. L. Corwin, Director. New York: Columbia University Press, 1952. Pp. 188. $3.50.

*Jim Crow.* Jesse Walter Dees, Jr., and James S. Hadley. Ann Arbor: Ann Arbor Publishers, 1951. Pp. 529.

*Informal groups and the community.* Hurley H. Doddy. New York: Bureau of Publications, Teachers College, Columbia University, 1952. Pp. 34.

*The development of executive talent.* M. Joseph Dooher, Editor. New York: American Management Association, 1952. Pp. 576. $6.75.

*Man into wolf.* Robert Eisler. New York: Philosophical Library, 1952. Pp. 286. $6.00.

*Personality measurement.* Leonard W. Ferguson. New York: McGraw-Hill Book Co., Inc., 1952. Pp. 457. $6.00.

*The annual survey of psychoanalysis.* John Frosch, editor. New York: International Universities Press, Inc., 1952. Pp. 556. $10.00.

*Behavior difficulties of children.* William Griffiths. Minneapolis: University of Minnesota Press, 1952. Pp. 116. $3.00.

*In search of self.* Arthur T. Jersild. New York: Bureau of Publications, Teachers College, Columbia University, 1952. Pp. 141. $2.75.

*Child adoption in the modern world.* Margaret Kornitzer. New York: Philosophical Library, 1952. Pp. 403. $4.50.

*Practical sales psychology.* Donald A. Laird and Eleanor C. Laird. New York: McGraw-Hill Book Co., Inc., 1952. Pp. 291. $4.00.

*Principles of human relations.* Norman R. F. Maier. New York: John Wiley and Sons, Inc., 1952. Pp. 474. $6.00.

*Conversation and communication.* Joost A. M. Meerloo. New York: International Universities Press, 1952. Pp. 245. $4.00.

*Psychiatry and medicine.* Leslie A. Osborn. New York: McGraw-Hill Book Co., Inc., 1952. Pp. 500. $7.50.

*The sensations, their functions, processes, and mechanisms.* Henri Pieron. New Haven: Yale University Press, 1952. Pp. 469. $6.00.

*Human nature, its development, variations and assessment.* John C. Raven. London: H. K. Lewis and Co., Ltd., 1952. Pp. 226. 12s. 6d.

*History of American psychology.* A. A. Roback. New York: Library Publishers, Inc., 1952. Pp. 426. $6.00.

*The dynamics of the counseling process.* Everett L. Shostrom and Lawrence M. Brammer. New York: McGraw-Hill Book Co., Inc., 1952. Pp. 213. $3.50.

*Recreation leadership.* Walter L. Stone and Charles G. Stone. New York: The William-Frederick Press, 1952. Pp. 81. $2.00.

*Readings in social psychology.* Revised edition. Guy E. Swanson, *et al.* New York: Henry Holt and Co., 1952. Pp. 680. $5.00.

*The psychology of thinking.* W. Edgar Vinacke. New York: McGraw-Hill Book Co., Inc., 1952. Pp. 392. $5.50.

*America's manpower crisis.* Robert A. Walker, Editor. Chicago: Public Administration Service, 1952. Pp. 191. $3.00.

*How to prepare and use job manuals.* Marguerite Holbrook Watson. New York: The William-Frederick Press, 1952. Pp. 38. $1.00.

*Society and personality disorders.* S. Kirson Weinberg. New York: Prentice-Hall, Inc., 1952. Pp. 544. $5.75.

*Group leadership.* Gunnar Westerlund. Stockholm: Sweden: Nordisk Rotogravyr, 1952. Pp. 257. 16 Swedish Crowns.

*Personnel principles and policies: modern manpower management.* Dale Yoder. New York: Prentice-Hall, Inc., 1952. Pp. 704.

*The British worker.* Ferdynand Zweig. Baltimore: Penguin Books, Inc., 1952. Pp. 240. $.65.

*Population and its distribution.* J. Walter Thompson Company. New York: McGraw-Hill Book Co., Inc., 1952. Pp. 428. $15.00.

*Personnel administration through supervisors.* A selected reading list. Pasadena: Industrial Relations Section, California Institute of Technology, 1952. Pp. 8.

*The first full report on*

# The Ames Demonstrations in Perception

*By WILLIAM ITTELSON*

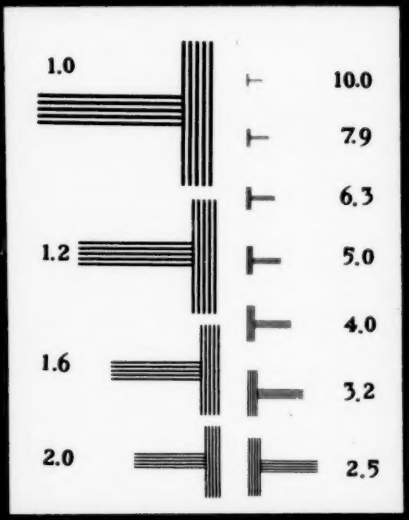
THE famous demonstrations in perception developed by Adelbert Ames of Hanover are here collected, pictured, and fully described for the first time. By ingenious photographs and drawings, the psychological effects are depicted and the construction of the apparatus shown. Since the theoretical interpretation of these demonstrations is still a matter of controversy, the treatment is factual throughout.

*104 pages, planographed, $4.00*

*Order from your bookstore,*

*PRINCETON UNIVERSITY PRESS*

# RESOLUTION CHART




100 MILLIMETERS

**INSTRUCTIONS** Resolution is expressed in terms of the lines per millimeter recorded by a particular film under specified conditions. Numerals in chart indicate the number of lines per millimeter in adjacent "T-shaped" groupings.

In microfilming, it is necessary to determine the reduction ratio and multiply the number of lines in the chart by this value to find the number of lines recorded by the film. As an aid in determining the reduction ratio, the line above is 100 millimeters in length. Measuring this line in the film image and dividing the length into 100 gives the reduction ratio. Example: the line is 20 mm. long in the film image, and $100/20 = 5$.

Examine "T-shaped" line groupings in the film with microscope, and note the number adjacent to finest lines recorded sharply and distinctly. Multiply this number by the reduction factor to obtain resolving power in lines per millimeter. Example: 7.9 group of lines is clearly recorded while lines in the 10.0 group are not distinctly separated. Reduction ratio is 5, and $7.9 \times 5 = 39.5$ lines per millimeter recorded satisfactorily. $10.0 \times 5 = 50$ lines per millimeter which are not recorded satisfactorily. Under the particular conditions, maximum resolution is between 39.5 and 50 lines per millimeter.

Resolution, as measured on the film, is a test of the entire photographic system, including lens, exposure, processing, and other factors. These rarely utilize maximum resolution of the film. Vibrations during exposure, lack of critical focus, and exposures yielding very dense negatives are to be avoided.

THIS PUBLICATION IS REPRODUCED BY AGREEMENT WITH THE COPYRIGHT OWNER. EXTENSIVE DUPLICATION OR RESALE WITHOUT PERMISSION IS PROHIBITED.